

Europe's need of learning modesty

European members of the European Space Agency are preparing for yet another crisis meeting. What they and their agency need is a proper sense of what can usefully be done in this novel field.

NEXT month's ministerial meeting of the European Space Agency (ESA) at The Hague on 9 and 10 November promises to be rowdy. The central issue (see page 660) is whether the thirteen member governments will endorse ESA's development programme for the coming decade, and thus implicitly agree that its budget should double (to roughly £2,000 million a year). Already, interested governments are declaring their positions. The British government is advertising its conviction that the budget is getting out of hand, perhaps hoping that advance publicity will persuade other governments to stand at its shoulder at The Hague. The government of France, by contrast, remains steadfast in the view that the sooner Europe is able to match grander powers such as the United States and Soviet Union in orbital adventures, the sooner Europe will succeed in other worthy causes, economic integration and improvement among them. Each government has a case. What is likely to be overlooked, in the manoeuvring for influence and allies, is that ESA is more than a device for emulating those who are bigger than Europe, that it has an estimable record both in basic and applied research and that these activities could be damaged seriously if the row about the budget itself gets out of hand.

The three most conspicuous items in ESA's new budget are centred around schemes for carrying people into space. One is a vehicle called Columbus, designed as an independent ancillary to the US space station. Another is the French conception of a manned space vehicle, called Hermes. The third is a heavy-lifting version of the successful Ariane satellite launcher which is necessary if Hermes is ever to leave the ground, but which would not be needed if there were no such plan. France is asking that all three projects should go ahead together. British ministers have been talking about schemes for altering the timetable for some of the components — the larger Ariane and Hermes could be postponed, for example, but at the risk of relying on US launching facilities to a greater degree than France considers wise.

Two questions should be in the minds of the ministers packing for the meeting at The Hague. First, how did it arise that they are now presented with such an unmanageable choice? The answer is that they (or their predecessors) agreed to support the developments so far at a meeting in Rome in 1982. Second, how has the issue of human orbital flight suddenly come to dominate ESA's plan of work? Answer: because nobody has bothered to ask tough questions of the enthusiasts for these escapades.

Questions

Sadly, European governments have drifted into the business of human spaceflight, or at least the preparations for it, without asking why. The government of France, with Hermes up its sleeve, has been so clear in its mind that the benefits of its preferred course are self-evident that it has hardly bothered to tell other people what they are. Other European governments, the British for example, have been sucked into the business by accident. It would have seemed unfriendly not to take up the offer by the United States of partnership in its space station, while the scheme to participate in a European adjunct to the venture seemed, five years ago, to be a way of winning the best of both worlds. Now the time has come either to pay the cost for

past indifference or to appear to renege.

The sensible course to follow, even at this late stage, is to go back to what should have been the beginning and to ask what this part of ESA's programme is meant to accomplish for Europe. One opinion is that the cause of European unity will be strengthened if European states are seen to be collaborating on an ambitious and successful programme, but that argument would be even more winsome if the common programme were, say, the creation of a successful European computer industry. Another is that Europe must be seen, for pride's sake, to be contributing to an historic international adventure, the exploration of the world out there by people. That opinion dubiously supposes that, if such a venture is eventually mounted, stretching over many centuries, it is nevertheless essential that participants should be there at the beginning. But the real Columbus was by no means the first of the great navigators, merely the one who won the most famous prize. So what of all the commercial operations that can be carried out by people riding in Earth satellites where gravitational forces are very small? There could be something in what enthusiasts for microgravity say about the preparation of small quantities of valuable materials in the absence of convection, but can it be enough to justify the investment of £2,000 million a year over a decade?

Courage

If only the gathering at The Hague had the courage to ask itself these questions, the result would almost certainly be a decision that no harm, but only good, would come from a postponement of the spaceflight parts of ESA's plans, at least for the time being. To be consistent, it would also be necessary to pull out of the US space station, already soured by the long row about the US Defense Department's long-declared interest in the venture, which seems to have taken some ESA governments by surprise a year ago. That, of course, might be the end for the time being of that enterprise as well, for NASA's recruitment of outside partnerships has always in part been regarded as a partial safeguard against budget cuts.

The conclusion, to which some will jump, that there will then be nothing left for ESA to do that is capable of yielding practical results is, of course, ridiculous, as the development of communications and broadcasting satellites shows. There may be limitations to the usefulness of photographs of the surface of the Earth by cameras mounted in satellites, but that lode is not yet worked out, nor is there a shortage of other things to be done. The British, after their experience at the weekend (hurricane-like winds in the south-east, torrential rains in the west) might have particular interest, just now, in efficient weather forecasting, for which instruments of sufficiently high resolution have not yet been designed. Questions about the Earth's climate, and the interaction between the solar wind and the Earth's atmosphere and magnetic field, although the oldest topics in the business, are still important and capable of yielding benefits. There should be enough interest in these and other geophysical exercises to keep the interests of researchers stimulated, and government ministers looking for a return on capital, until the future is much clearer than it will be by 9 November. □



Does M'Bow bow out?

Management is changing at UNESCO. Spain's Frederic Mayor needs a persuasive programme.

THE Executive Committee of UNESCO (the United Nations Educational, Scientific and Cultural Organisation) deserves credit for having walked through the fire of Mr Amadou M'Bow's displeasure and the anger of the African governments supporting his bid for a third term as director general. By the weekend, after making it a condition for one last adjournment of its complicated voting procedure that there "shall be no more adjournments", the committee decided to nominate Dr Frederic Mayor, the Spanish biochemist, as M'Bow's successor. Until Mayor's election has been confirmed by UNESCO's general conference next month, it would be wrong to suppose that UNESCO is now back on an even keel. There is at least one other occasion when the Executive Committee's recommendation has been overturned. On this occasion, Mr M'Bow's constituents will be lobbying hard until the last minute, perhaps even afterwards. But on the face of things, reason has emerged on top.

This journal, although it supported another candidate (see *Nature* 329, 472; 1987) offers Mayor warm congratulations alloyed with the comiserations appropriate to the task ahead of him. One of his qualifications for the job is that he is a scientist — a biochemist — by training. Another is that he is deft at persuading people that what he wants them to do is in their best interests. During his spell as Minister of Education in the post-Franco government of Spain, he used this skill to found what may still be Spain's best-known laboratory, the Centro Biología Molecular at Madrid. That, more than his previous experience (during M'Bow's first term as director general) of UNESCO's problems from the inside, is what the demoralized organization in Paris will most value.

In reality, the task ahead is not what it is usually likened to, that of the Aугean stable, but the more difficult yet more interesting task of giving a rudderless ship a sense of direction. The first need is to escape from UNESCO's preoccupation over the past several years with its own administration. If Mayor can also devise a programme of work uniting member governments in the belief that UNESCO is embarked on tasks that are at once useful and feasible, he may even win the greatest prize of all: the state of grace in which members are so interested in what is going on that they have no interest in using UNESCO as a stalking horse for their political objectives.

There are also some preliminary points that need to be established. UNESCO is an international organization involving both rich and poor countries, but it does not follow that UNESCO must behave as if it were just another aid organization. Indeed, its budget is too small to make much of a dent in that role (which is not the same as saying that there is no part for it to play in the process of development, or that no attention should be given to the special problems of the poor).

Second, UNESCO's much maligned staff is not the uniformly malevolent influence on affairs it is made out to be. Many of its members are, by contrast, able and knowledgeable people whose frustration in recent years has been that maladministration has prevented them working effectively. Their demoralization can be lifted easily; the more difficult task will be to win their consent to the proposition that members of the UNESCO staff are not by definition scholars, but people whose skills are (or should be) in the administration of projects usually conceived by others. The new administration will need to welcome more readily than its predecessors advice from independent scholars. (To his credit, M'Bow did at one stage set up an advisory committee on scientific projects.)

Finally, the new administration must not be thrown off course by the likelihood — indeed, the certainty — that much of the

work it may attempt will raise political issues. UNESCO's successful sponsorship of the negotiations leading to the Paris amendment of the Berne Copyright Convention, for example, entailed a political compromise between governments with powerful publishing industries and those without them. More contentious issues would arise if UNESCO sought (as legitimately it might) to define how and when legitimate scholars and researchers should be free to move about the surface of the Earth. The new UNESCO should welcome work whose consequences are so important that some of its members regard them as political, for that will be a sign that it is ploughing where it may be profitable to sow seed. What it must somehow avoid is the politicization of which the United States (and others) have rightly complained, and which has blocks of countries voting in unison.

What projects, in the circumstances, to tackle? Mayor's difficulty (if he eventually gets the job) will be that there are too many useful things to do, not too few. How best to extend the academic computer networks now covering the rich part of the world to the places still out in the cold? What about an equitable convention on the use of computer software? What is to be done with and about the mountains of geophysical data being accumulated by Earth satellites, or by the nucleotide-sequencing banks? Is there yet an agreed basis for an international convention of the procedures that should be followed in genetic manipulation? To what extent does a practising scientist have an allegiance to science rather than to his or her employer? Mayor does not have to tackle all these issues at once; he will, in any case, have to demonstrate that he is more than a mere biochemist willing to throw culture out of the window. But at some stage, say a year from now, he will need an important theme. Some people say it should be science education: science education for all, everywhere. It is a challenging idea. But it could be made to work. □

Bubble goes Bang!

Those seeking a scapegoat for the past week's stock market crash had better blame the President.

PEOPLE with an interest in demonstrating that water will run uphill are usually adept at rigging the topography, but the president of the United States for the past seven years is made of simpler stuff. He believes that taxes imposed by governments (of one of which he is the head) are intrinsically bad, he knows in his bones that he is right, and he has steadfastly resisted the notion that taxes in the United States should be increased to the point at which his treasurer, Mr James Baker, would not have to go every week to Wall Street to parley people into lending their surplus cash to the government of the United States.

As always, the consequences of irrationally devised policies have been paradoxical. Mr Baker's enticements have persuaded his fellow countrymen to lend to the US government, allowing others (from Japan and even Western Europe) to snap up the investments disposed of in the pursuit of relatively higher income. Mr Baker has even spent much of the past week complaining at West Germany's increase of interest rates two weeks ago, forgetting that Bonn (or, rather, Frankfurt) has merely been fighting for its share of liquidity. The good news, now, is that the US Congress has given itself until mid-November to settle the budget for the present year; Mr Reagan could clean up much of the mess he has created by asking the Congress to raise a tax — a few cents on gasoline would do the trick.

But that will probably not happen, if only from the fear that next year's election would be influenced, in which case there will be a real crash. *Nature's* concern in this matter stems simply from the circumstance that most of its readers contribute to now-devalued pension funds. □

UNESCO finally elects a new director general

- Executive council elects Federico Mayor
- M'Bow support fails at final vote

Paris & Washington

UNESCO (the United Nations Educational, Scientific and Cultural Organisation) has nominated Federico Mayor Zaragoza, a Spanish molecular biologist, as its new director general. A majority consensus among the 50 delegates of the executive council was reached last Saturday evening (17 October) after the full five rounds of voting, lasting 10 days.

During the first four rounds, the incumbent director general, the controversial Amadou Mahtar M'Bow, of Senegal, remained in the lead, despite efforts by most Western member states to dislodge him (see *Nature* 329, 472; 1987).

M'Bow, who was encouraged to stand for a third term by UNESCO's third-world members, reluctantly agreed to quit the contest on Saturday.

The elections took an unexpected turn on 9 October when Sahabzada Yaquub Khan, the Pakistan candidate favoured by Western Europe, Canada and Japan, resigned from the contest after being runner-up in the first three rounds. Hopes by the anti-M'Bow lobby that Khan's votes would go to Mayor, giving him an overall majority, were upset when France (which retains close links with several African countries) decided to back M'Bow. It was only in the final, fifth round that France cast its vote for Mayor.

Squibb boosts Oxford pharmacology lab

London

THE US drug house E.R. Squibb is giving £20 million to the department of pharmacology at the University of Oxford to support established research into the chemical basis of disorders of the central nervous system. Richard Furlaud, chairman of Squibb, says the company is attracted by the quality of neuroscience research in Oxford, and that it recognizes the need for a rational basis for treating diseases such as Alzheimer's and schizophrenia.

Half of the money will be used to rehouse the department of pharmacology: the balance will support selected research projects over the next seven years, with a possible extension for a further five years. Researchers will not be unexpected to produce marketable products, but Squibb will retain the rights on discoveries with patent royalties going to the university.

Jennifer Altman

M'Bow's alleged mismanagement of funds and anti-Western bias led the United States then Britain and Singapore to withdraw from UNESCO in 1984 and 1985.

Mayor is a biochemist by training and carried out research at the University of



Federico Mayor, the new director general.

Oxford in the laboratory of Sir Hans Krebs. He is a strong supporter of funding for basic research but also a pragmatist who believes that politicians need to be taught what science can do for society. He became one of Spain's youngest professors when he was given a chair at the University of Granada at the age of 29. More recently, he provided the impetus, along with Nobel prizewinner Severo Ochoa, to set up Spain's main research institute, the Centre for Molecular Biology, at the Autonomous University of Madrid. Although he served briefly as Minister of Education in the 1970s, he has not been much favoured by the present socialist government. But he is known as a man with formidable political connections and is a friend of the King of Spain.

Speaking at a news conference in Paris on Saturday, Mayor said that he intends to restructure UNESCO and to encourage Britain, the United States and Singapore to rejoin the organization. Britain, however, has said that it would be at least two years before it would consider rejoining and, with the United States, it would expect to see major changes to the way UNESCO is run.

The executive council's nomination of Mr Mayor has still to be ratified at a full assembly of the 158 UNESCO members states on 14 November.

Peter Coles & Alun Anderson

Berlin's second Academy of Sciences opens

Munich

LEIBNIZ, the founder of Berlin's first Academy of Science in 1700, is a tough act to follow, but West Berlin's conservative city fathers have decided to try. They have founded the "Academy of Sciences at Berlin", an alternative to the Leibniz academy now located in East Berlin. The new academy opened to mixed reviews at its founding in the former Italian Embassy in Berlin's Tiergarten on 11 October.

Opposition politicians, such as Hans Kremendahl of the Social Democrats (SPD), consider the new academy nothing more than a 'debating club' or, worse still, a conservative think-tank, which the SPD would redesign from the ground up if given the chance.

Academy scientific director Helmut Meier refutes the SPD charges, saying that the academy is the more capable because its members are scattered "from Freiburg to Harvard". The new academy, which will ultimately expand to 60 members, is no "old men's discussion circle", he said, "occupied with a debate over whether June bugs have six legs or seven". Nevertheless, Meier agreed that the orientation of the academy "does not exactly fit SPD educational policy".

The academy, which will have a budget for 1988 of DM6 million, begins its work with 30 members in fields ranging from gerontology to energy. Its most innovative aspect is its structure, which calls for seven 'working groups' consisting of members both inside and outside West Berlin and West Germany, to tackle thorny conceptual problems. The groups themselves will have no facilities, but rather will rely upon the home institutes and research grants of members. The academy hopes, by pulling people together from different fields, to break down traditional barriers.

Initial topic areas include "the long-term chances of solar energy" and "ageing and society" as well as an analysis of the interactions between mathematics and physics in the field of string theory. The only topic with a specifically German focus will assess the impact on German science of the 1933 exodus of researchers from Berlin caused by the Nazis.

Until now, West Germany has had no national Academy of Sciences, although there are four regional academies and one "Academy of Arts, Sciences and Literature" in Mainz. The academy founded by Leibniz as the "Prussian Academy of Arts and Sciences" is now called the "Academy of Sciences of the German Democratic Republic", with a budget of 1,000 million East German marks.

Steven Dickman

Agencies anxious as US Congress fixes deadline for 1988 budget

Washington

US SCIENCE agencies are watching anxiously as Congress enters the home stretch of the 1988 budget process. Although the financial year began officially on 1 October, Congress has given itself until mid-November to make all its outstanding spending decisions for 1988.

Many agencies are enjoying a period of cautious optimism as the final decisions draw near. But there remains the threat that last-minute compromises to balance the books could radically alter the financial landscape. And if that were not enough to cause sleepless nights in budget offices throughout Washington, there is also the spectre of automatic cuts brought on by the modified Gramm–Rudman–Hollings deficit reduction act.

Both the House of Representatives and the Senate have now completed work on the National Science Foundation (NSF) budget, but differences will have to be worked out by a joint House–Senate conference. The Senate appeared ready to make significant cuts in NSF's budget (see *Nature* 328, 749; 1987) when, at the last moment, extra money was found.

The Senate appropriation for NSF now stands at \$1,867 million, just slightly under the administration's request. The House appropriation is \$1,796 million. The House version cuts NSF research money, but increases funds for science education. The Senate reduces the amount for NSF Antarctic programmes.

The National Institutes of Health (NIH) have received their accustomed increase from Congress, with the result that the budget is well ahead of the Reagan administration's request. Excluding money for research on AIDS, the House would provide NIH with \$6,563.9 million and the Senate \$6,388.2 million. For AIDS research, the House adds \$472.4 million and the Senate \$487.9 million.

According to Katherine Bick, NIH deputy director for extramural research and training, the amounts proposed should allow between 6,200 and 6,400 new research awards to be made, although the usual trimming of individual financial awards may be slightly larger than usual.

Both House and Senate bills contain money for the National Institute of General Medical Sciences specifically for the project to construct a genetic and physical map of the human genome, but the House version provides \$30 million compared with \$6 million in the Senate.

For the National Aeronautics and Space Administration (NASA), the biggest question mark is the space station. The Senate bill provides some \$200 million less than the requested \$767 million.

Should the Senate version prevail, the programme would suffer a severe setback. US negotiators are still trying to nail down agreements with Japan, Canada and the European Space Agency on participation in the project. Those agreements must be reached soon, or hardware development and production will be delayed.

It is still unclear how the final budget plans will be translated into law, as there is not enough time for conferences to resolve all the differences between the various versions of the appropriations bills. All the appropriations could be incorporated into a single, giant spending

bill. But putting all the eggs in one basket often pits Congress against the president in a test of wills, as the president has vetoed budgets he is not happy with.

Both House and Senate are also looking at ways to raise taxes to provide the revenue that will be needed to keep the deficit from growing even faster than at present, but the White House has adamantly opposed increased taxes as a way of shrinking the deficit. The revised Gramm–Rudman law passed last month forces spending to be curtailed, but the new targets are not as stringent as those in an earlier version. The law calls for a deficit cap of \$144,000 million for 1988, but restricts automatic cuts to \$23,000 million. If the automatic cuts take place, agency budgets could be reduced from 3 to 8 per cent.

Joseph Palca

Tough bargaining ahead on European space programmes

London & Paris

THERE are increasing indications that next month's ministerial meeting of the European Space Agency (ESA) in The Hague will not be the formality for which officials are hoping. Last week, the British government stated for the first time publicly that it would be seeking to persuade the meeting to adopt a less ambitious programme for the next decade — current plans are for an integrated package of elements comprising a heavy-lift rocket launcher (Ariane V), a space platform (Columbus) and a shuttle vehicle (Hermes). Adoption of the proposals would mean an increase

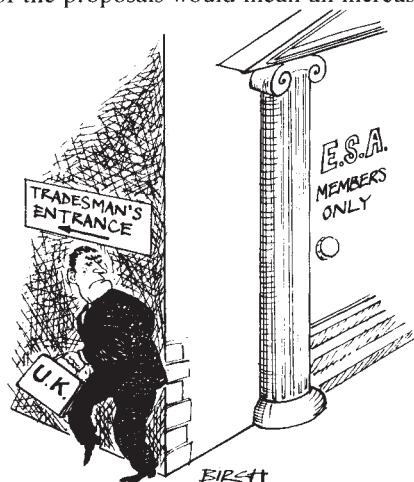
think ESA's proposals are expensive and not all of the present proposals need to be decided now. We believe they should re-think their proposals and see if they can instil a more commercial slant into their operations."

Furthermore, Butcher claimed that the government's concern about "the trend in ESA towards widening development objectives, with apparently diminishing regard for commercial realities" is not Britain's alone. "We are saying publicly what other governments are saying privately." There are indications that West Germany, for one, would like to see the ESA meeting postpone adoption of the development phase of Hermes until the next decade. ESA officials in Paris are reluctant to comment on the British government's position. The hope within ESA is that any doubtful governments will be convinced at the meeting of the need for the expanded programme. The more optimistic recall the spontaneous enthusiasm at the last ministerial meeting in 1982 when the design phases of the three big programmes were adopted.

Expanding investment in space, particularly by the Soviet Union but also by Japan and the United States, will be cited as one of the main reasons for the need for European cohesion in space. One feeling within Britain is that Clarke and Butcher have timed their announcements carefully to serve as a shot across ESA's bows that tough negotiations lie ahead.

Meanwhile, agreement has still not been reached on collaboration between ESA and NASA on the international space station (see above). Intervention by the US Department of Defense has made it difficult for governments to agree on terms for the collaboration.

Simon Hadlington & Peter Coles



in ESA's £1,000 million budget from 70 to 100 per cent.

Britain has already said that it will not increase its contribution, with Kenneth Clarke, the minister responsible for Britain's space programme, describing ESA as a hugely expensive club with over-ambitious programmes. Last week, Clarke's number two at the Department of Trade and Industry, John Butcher, told space industrialists in Brighton that "we

Changes ahead for US policy on biotechnology regulations

Washington

THE White House Office of Science and Technology Policy (OSTP) is at the centre of a policy struggle over the control of federal oversight of biotechnology.

There are signs that OSTP is seeking to wrest control of oversight from the Biotechnology Science Coordinating Committee (BSCC), and to expand its charter to include jurisdiction of federal efforts to map and sequence the human genome and human gene therapy.

In a bizarre twist, press reports surfaced

last week that BSCC chairman David Kingsbury is under investigation by the Justice Department for possible conflict of interest arising from his association with a British biotechnology company.

BSCC was established by former White House science adviser George Keyworth II in 1985 as part of the Federal Coordinating Council for Science, Engineering and Technology. It is made up of senior officials of the federal agencies involved in biotechnology research and product regulation.

What future now for Japanese biotechnology research?

Osaka

WHERE is biotechnology headed in Japan? The BIO87 conference held in Osaka last week gives some indication.

The early days, when giant chemical, distilling and pharmaceutical companies paid high prices to license frontier technology from small US venture concerns, are now long gone. So is the fad for interferon which preceded the rush to produce tumour necrosis factor and colony stimulating factor.

According to Mitsuru Miyata, editor of *Nikkei Biotech*, who spoke at the conference, there is a growing belief that the future of biotechnology lies not in the mass-production of proteins but in gaining knowledge of the structures and functions of the molecules, for example, through protein engineering. And about 20 biotechnology companies, including the 'big five' (Kyowa Hakko Kogyo, Takeda Chemical Industries, Toray Industries, Mitsubishi Chemical Industries and Ajinomoto) and Toa Nenryo Kogyo, an oil-refining company, will open the Protein Engineering Research Institute in Osaka next autumn (see *Nature* 325, 651; 1987).

But such research and development will have a very long lead time, probably of more than a decade, and in the meantime companies are promoting small-scale enterprises using existing technology. Already cell-fusion technology is aiding the production of *sake* and orchids are being propagated using tissue culture. Japan's 47 prefectures have more than 100 biotechnology promotion centres and biotechnology companies are establishing links with them and with agricultural cooperatives to disseminate knowledge.

Plant biotechnology featured prominently in the conference. Building on successes in growing rice from protoplasts (*Nature* 321, 639; 1986) Dr K. Shimamoto of Plantech Research Institute

has succeeded in growing 250 hybrid plants created by electro-fusion of protoplasts of cultivated and wild rice. All have grown to maturity and some have started to flower. The development will give rise to the virtually untapped genetic variability in the ten species of wild rice.

The Ministry of Agriculture, Forestry and Fisheries is also showing keen interest in rice biotechnology, and 70 million (\$500,000) has been allocated for a feasibility study for full sequencing of the chromosomal DNA of rice. But, according to Miyata, the ministry sees many technical difficulties ahead and may favour a simpler mapping project. The ministry has already paid out ¥290 million (\$2 million) to the United States for the genome map of the tomato plant, and this will be used with conventional technology to create new nutritious tomatoes.

Japan has gone a long way towards catching up the West in biotechnology. Technology is now beginning to flow the other way; for example, Genentech has signed a five-year agreement with Mitsubishi Chemical Industries for rights to Mitsubishi's biotech pharmaceuticals.

So will the future see trade disputes erupting between Japan and the United States over biotechnology? Ronald Cape, chairman of Cetus corporation, hopes not. He says that the lead times between invention and marketing of biotechnology products is often long compared with other high-technology goods, partly because of problems over patenting and government environmental and health regulations. And he believes this delay will provide opportunities for Japan-US cooperation in research and development. And as a step in that direction he proposed a "Harmony Fellowship" scheme for international exchange of late-postdoctoral fellows with initial funding from the private sector.

David Swinbanks

Kingsbury, assistant director of the National Science Foundation (NSF), and director of the division of biological, behavioural and social sciences, has chaired BSCC since its inception.

In June 1986, the working group promulgated the coordinated framework for regulation of biotechnology, an attempt to rationalize the interlocking jurisdictions of federal agencies and to establish uniform policy (see *Nature* 321, 458; 1987).

Kingsbury and BSCC have received mixed reviews for their efforts. The biotechnology industry has been generally pleased, but critics say the framework has failed to achieve its goal of coordination, and has left far too many loose ends, allowing potentially risky experiments to slip through regulatory cracks.

Kingsbury has been an energetic champion of the framework and BSCC. He also has sought to expand the role of the Domestic Policy Council working group on biotechnology. Earlier this year he established a task force that was intended to oversee federal involvement to map and sequence the human genome (see *Nature* 325, 651; 1987).

But BSCC's charter was written with a sunset clause that puts it out of existence at the end of this month unless action is taken to renew it. OSTP has been seeking input from membership of BSCC, as well as other government agencies, about how a new body should be constituted. OSTP spokesperson Mary English says some decisions about the future of BSCC have been made, but OSTP intends to brief federal agencies before releasing details of its plans.

Kingsbury's role in the debate has been muted by the Justice Department investigation. He is accused of maintaining an association with Porton International plc., a British biotechnology company with offices in the United States.

Porton claims that Kingsbury was a scientific advisor to Porton subsidiary IGB Products, Ltd. The House of Representatives Committee on Space, Science and Technology got wind of the potential conflict last April, and asked the Reagan Administration to look into it. An investigation by Kingsbury's own agency was delivered to NSF Director Erich Bloch, who decided that "no immediate administrative action would be appropriate".

News of the investigation appeared in print for the first time last week, just as OSTP was planning to make its final decisions concerning the future of BSCC.

Kingsbury admits that he worked as a consultant to Porton, but says he suspended this relationship when he joined NSF. He also admits receiving stock in IGB Products while at NSF as part of his consulting arrangement, but says he returned the stock to the company. Porton officials confirm that Kingsbury has no current connection with the company. Joseph Palca

US pressures Japan over imbalance in basic research

Washington

THE little-known US-Japan Science and Technology Agreement may be due for a rise in status. The agreement, a vague umbrella accord first signed in 1980, has never been the stuff of which headlines are made. But negotiations for its renewal are under way with the US side eager to make the agreement incorporate concrete measures to reverse the flow of science and technology from the United States to Japan.

The agreement, originally signed by President Carter and Prime Minister Ohira, is set to expire on 31 October. To White House science adviser William Graham that seemed to provide a timely opportunity to try to make some changes in the US-Japan relationship. But change is not going to come quickly; the first round of negotiations ended on Monday with the dead-line for expiry put back three months in order to give more time for negotiations.

Pressure for change has been building on both sides of the Pacific. In the United States there has been some resentment at the numbers of Japanese researchers who study at top US universities and research institutes and then return home to put their knowledge to work in commercial products. In 1986 there were 323 Japanese researchers at the National Institutes of Health; all had won their places in open competition but many came from biotechnology companies which will one day compete with their opposite numbers in the United States. Particularly galling is that 90 per cent of the visiting Japanese are supported by US government grants.

In return, only a handful of US scientists study in Japan. But that does not mean Japan is content with the situation. The obvious weakness of Japanese basic research is an embarrassment and stimulates endless national self-examination — but little action. Even the much touted Human Frontiers Programme, intended to put Japan at the helm of a massive international programme of basic research, seems destined to follow Prime Minister Yasuhiro Nakasone into retirement.

In evidence before a congressional hearing on the eve of the negotiations, both Graham and John Negroponte, who is responsible for international scientific affairs at the State Department, stressed that the administration is hoping that Japan will "act aggressively" to find ways to achieve "balance and reciprocity" in basic scientific and technological exchange. Specific proposals have been worked out and were incorporated in a diplomatic note transmitted to Japan in August. The proposals, which remain

confidential in their details, are not expected to win ready acceptance from Japan.

"Reciprocity", rather than equality, has become the aim as negotiators recognize the huge differences in the two nations' cultures. The US strength is in basic research, carried out in universities and national laboratories which are open to all. Japanese universities and government institutes support less research. And while in theory they are open — recent reforms have even made it technically possible for foreigners to hold permanent posts in Japanese national universities — it is very hard even for Japanese researchers to move from one laboratory to another. Postdoctoral fellowships, which give some independence to young researchers, are only just being introduced. But more places for US researchers at Japan's research institutes should be negotiable.

Japan's strength in applied research is what the US administration would like to tap most of all. But access to applied research in exchange for the continuing Japanese presence in US basic research laboratories will be difficult to negotiate, however attractive it may seem. Most of Japan's applied research is carried out in company laboratories where the government cannot easily demand change, although it can influence programmes where it provides part of the funds.

Although several witnesses at the congressional hearing clearly felt angered by Japanese exploitation of US research, others saw much of the blame lying in the United States. That 70 per cent of Japan's technology imports have come from the United States is seen by some as symbolic of US failure to capitalize on its own basic research. And the biggest barrier to access to Japanese laboratories may not be the government but the language. There is no evidence that US researchers are being turned away from Japanese laboratories in any numbers. Nevertheless, the US administration view appears to be that as Japan profits from its relationship with the United States it has to find some way to set matters right.

With the next round of negotiations scheduled to take place in Tokyo in December, it is clear that quick results are not expected. But the Japanese side has taken the message that the imbalance in basic research is now an important issue in US-Japan relations. Japan cannot afford to be complacent. Waiting in the wings are those who would like to see all Japanese researchers banned from the US laboratories, or at the very least see part of the US research budget charged to the Japanese government as a condition of their continuing presence. **Alun Anderson**

Chemistry prize for makers of macromolecules

Washington

THIS year's Nobel prize in chemistry was almost awarded to Donald O. Cram, a Los Angeles carpet cleaner who happens to have a bachelor's degree in chemistry. The Nobel committee fortunately retrieved its mistake and was able to make the award to Donald J. Cram of the University of California at Los Angeles, Jean-Marie Lehn of Louis Pasteur University, Strasbourg, and the Collège de France, Paris, and Charles J. Pedersen, now retired, of the du Pont company in Wilmington, Delaware.

The winners are between them responsible for creating 'supramolecular' or 'guest-host' chemistry, in which the shape of the participating molecules is fundamental to the progress of reactions. Pedersen did much of the early work, and his original goal was to find a way of controlling the action of a vanadium catalyst in a scheme for producing a synthetic rubber. This never came to fruition, but Pedersen discovered in the course of his efforts cyclic molecules — crown ethers — which could accommodate in a central position in their structure a metallic ion. Cram and Lehn extended this work, creating crown ethers which attach to amino acids and then more complex molecules which, by joining in a specific way with simple organic molecules such as acetylcholine, mimic the chemical reactions of enzymes.

The experience gained by Lehn and Cram in particular has led to more quantitative detailed understanding of how molecular properties generate selectivity in reactions. For the time being, their work is 'basic research at its most pure', but in the long term there are obvious indications of significance for biochemistry. The selective transport of ions across biological membranes and the mimicking of enzyme chemistry are cited by the Nobel committee as two areas where guest-host chemistry can not only improve understanding but also engender applications, such as the extraction of toxic metals from living organisms.

The three winners have never worked together, and Cram likes to describe them as 'colleague-competitors'. According to Harvey Lubo, a chemist in Cram's department, the research recognized in this year's prize is the result of various different lines of investigation coming together, rather than an outgrowth from a central idea. But the awarding of a Nobel prize is a sign that guest-host chemistry is now a legitimate speciality, with methods and goals of its own. **David Lindley**

Swift Nobel physics prize for superconductivity researchers

London

A YEAR of unprecedented excitement in solid-state physics has been capped by the award of this year's Nobel prize in physics to the two scientists who started it all off. J. Georg Bednorz, 37, and K. Alex Müller, 60, both of the IBM Zürich Research Laboratory in Rüschlikon, Switzerland, have been cited for "the discovery of superconductivity in ceramic materials". Their report, published just 13 months ago, of superconductivity at 35 K (-238°C) in a mixed oxide of lanthanum, barium and copper triggered an explosion of research activity that has already led to superconductors that work at liquid-nitrogen temperature, with the promise of room-temperature superconductors just around the corner.

Bednorz and Müller had been looking for superconductivity in mixed oxides since 1983. The superconductors with the highest transition temperatures (T_c s) then known were alloys of metals, with the record standing at 23.3 K for Nb₃Ge. But this record had been set in 1973, and ten years of experimentation with other intermetallic alloys had failed to increase T_c . Müller, who had worked on insulating oxides in other contexts, focused instead on the known oxide superconductors, chief among which was a mixed oxide of barium, lead and bismuth, BaPb_{1-x}Bi_xO₃, with a T_c of 13 K and a crystal structure belonging to the class called perovskite.

Using crystal-chemical arguments to narrow the search, Müller decided to concentrate on perovskites containing nickel or copper, and enlisted the help of Bednorz, a skilled experimentalist. After two years without success, the breakthrough resulted from reading a paper by Bernard Raveau and colleagues, at the University of Caen, which described the synthesis of a lanthanum-barium-copper oxide perovskite with copper in a mixed-valence state. Bednorz and Müller set out to synthesize this compound, and here luck intervened. Bednorz and Müller's synthesis procedure was subtly different from that used by the Caen group, and led to a mixture of phases, instead of the pure phase they had been trying to make. One of the 'rogue' phases — also a mixed-valence oxide with a perovskite-related structure — turned out to be the 35-K superconductor.

The events that followed are now well known: the confirmation of Bednorz and Müller's results by Shoji Tanaka's group in Tokyo, followed rapidly by groups at the University of Houston, Bell Laboratories, Bell Communications Research and the Chinese Academy of Sciences; the race to even higher T_c s, led by Paul Chu at Houston and collaborators, culminating in the discovery by Chu's group (and perhaps simultaneously by Z.X. Zhao and colleagues in China) of superconductivity at 90 K (above liquid-nitrogen temperature)

An Abitur agreement

Hamburg

AFTER more than two years of discussions, the ministers in charge of culture and education of the West German *Länder* (regions) have agreed on a uniform *Abitur*, the final examination after nine years in the gymnasium (secondary school). The crucial point is that from 1989 each *Länd* will accept the *Abitur* of another *Länd*. The Social Democratic model of *Kolleg-Schulen* (in Nordrhein-Westfalen), which includes vocational education, was also accepted by Christian Democratic ministers. Everyone at a gymnasium must now study two of three subjects — German, a natural science (including mathematics) and one foreign language — through the *Oberstufe* and as an *Abitur* examination.

Jürgen Neffe

in the yttrium-barium-copper oxide system.

Chu had been widely tipped as a strong candidate for a share of the Nobel prize, and there are those in the field who feel that he deserved the honour. (One physicist pointed out that, owing to the time taken to publish their results, Bednorz and Müller had a nine-month start on Chu, and yet they did not find the 90-K superconductor.) Others argue that Chu's breaking of the liquid-nitrogen barrier was more of a technological breakthrough than a scientific advance (it is not yet known whether the physics of the 90-K superconductors differs from that of their 35-K relations), and that its importance will depend on the rate of progress to still higher temperatures.

But the Nobel committee's choice may have been governed by more practical considerations: by giving the prize to Bednorz and Müller alone they avoided any controversy over priority between the Chinese and Houston groups; also, Chu's paper describing the 90-K superconductor did not appear until more than a month after the 1 February deadline for Nobel nominations. Bednorz and Müller themselves are said to have just slipped in before the deadline, the result being a rare approach to fulfilment of the words of Alfred Nobel's will, which states that prizes shall be given to those who, "during the past year, shall have conferred the greatest benefit on mankind".

When the congratulations are passed round in Rüschlikon this week, some will certainly be due to the laboratory itself, for the second Nobel prize in as many years. Like last year's winners of the physics prize, Gerd Binnig and Heinrich Rohrer (inventors of the scanning tunneling microscope), Alex Müller is an IBM Fellow. IBM's policy of allowing fellows to pursue projects of their own choosing for five-year periods thus appears to have paid off handsomely, at least in prestige.

Laura Garwin



This year's line-up of Nobel laureates in the sciences. Top row: immunologist Susuma Tonegawa, winner of the prize in physiology/medicine (see *Nature* 329, 570; 1987), physicists Alex Müller and Georg Bednorz (see above). Bottom row: the chemists Donald J. Cram, Charles J. Pederson and Jean-Marie Lehn (see left).

French AIDS campaign criticized as too soft by volunteer groups

Paris

THE French Ministry of Health launched the third part of its information campaign against AIDS (acquired immune deficiency syndrome) — a 20-second video clip — on all five channels of French television last week. The film, which is said to have cost the government FF7 million (\$1.13 million), will be shown 90 times and shows a heterosexual couple intending to use a condom before having sex. The message of the film is "it won't pass on through us".

The film — and the government's AIDS

UK barn owl decline cause for alarm

London

CHANGES in weather patterns and farming methods have combined to cause an alarming decline in the population of the barn owl (*Tyto alba*) in Britain over the past 50 years. According to a census in 1932, there were then an estimated 12,000 pairs in



The barn owl — descending towards extinction?

England and Wales. A study carried out between 1982 and 1985 by Dr Colin Shawyer, of the Hawk Trust, published this week*, shows that the 1932 figure has dwindled by nearly 70 per cent, to 3,750.

The decline has been caused mostly by a reduction in the availability of food, impairing the birds' ability to breed successfully or to feed their young. Prolonged periods of snow, far more prevalent during recent years, have limited the owls' ability to find their staple diet of small rodents.

Modern farming methods have also made food harder to come by. The trend could be reversed by the immediate implementation of "effective conservation measures", including recreating grassland, says Shawyer. Simon Hadlington

prevention campaign in general — has already been criticized as being too soft and for missing high-risk populations altogether. So far this year, the government has printed 13 million leaflets on AIDS, 9 million of which have been delivered to telephone subscribers. The remaining 4 million are to be delivered to doctors' surgeries, the armed forces, universities and secondary schools. No special measures have been taken by the government to make sure that those most at risk — sexually active homosexuals, drug addicts and prostitutes — receive information on AIDS prevention.

The *leitmotif* of government advertising and politicians' speeches is that AIDS will be stopped only by giving money for medical research. The early success of the Pasteur Institute in isolating the virus causing AIDS made it possible for the search for an AIDS vaccine to become not only a "cause of national importance", but an occasion for national pride. This pride, according to critics, should not obscure the need for prevention. With 1,964 known cases of AIDS (the latest official figures) and over 150,000 carriers of HIV (human immunodeficiency virus) antibodies, France has the highest incidence of AIDS per head in Europe.

In response to the perceived weaknesses of the government information campaign, several voluntary AIDS information organizations have started in France. The Association pour la Prévention du SIDA has a 24-hour telephone-answering service, although the recorded message only gives further numbers to ring. One of these numbers belongs to another group, called AIDS, which gives the times and numbers of telephone advice 'surgeries' throughout the country.

The most dynamic and perhaps effective organization is AJIS (Association de Jeunes pour l'Information sur le SIDA), a group of young people who have launched a 'no-holds barred' information campaign, using the FF500,000 (\$81,000) donated by pop-singer Madonna after her Paris concert this year. This group has already made a film on safer sex and has distributed three different leaflets to secondary schools, in editions of 450,000. The last edition contained a free condom, with clear instructions for its use.

This week, FF3.8 million (\$613,000) was raised for the AIDS organization and for medical research by an auction of paintings. The French left-wing daily, *Liberation*, also carried a half-page open letter signed by 400 doctors belonging to SIDAventure, dedicated to "the development of a network of vigilance to ensure that medical opinion is not used for racist

or segregationist purposes". The letter was an open rejection of opinions expressed by Dr Bachelot, a member of the extreme right-wing National Front party, who had suggested imprisoning all AIDS patients in secluded "SIDAtoiria".

• AN international ethics committee on AIDS is to be set up, following a suggestion mooted at the Venice summit in June. French President François Mitterrand has named France's representatives: Professor Jacques Ruffie, of the Collège de France; Alain Pompidou, technical adviser to the health ministry; and Dr Willy Rozenbaum, of the Claude Bernard hospital in Paris. President Reagan has selected as the US representatives Eugene Mayberry, of the Mayo Clinic at Rochester and president of the US AIDS commission; Charles McCarthy of the National Institutes of Health; and Leroy Walters of the Kennedy Institute at Georgetown University.

The French minister for research and higher education, Jacques Valade, has released funds set aside for AIDS research (see *Nature* 328, 191; 1987). FF20 million will go to industry as venture capital for the search for a vaccine. FF60 million will be divided between medical and science research institutes. Peter Coles

Raman controversy puts award on ice

New Delhi

THE Indian National Science Academy (INSA) has ended an embarrassing two-year controversy by deciding that offers of its coveted research awards will in future be withdrawn if those to whom they are made do not begin research within a year.

The decision arises from the award in January 1986 of the Sir C.V. Raman research professorship to Professor M.G.K. Menon, one of India's leading science administrators and a well-known cosmic-ray physicist who is now also the president of the International Council of Scientific Unions.

During 1985, Menon had apparently let it be known that he was thinking of quitting his job with the Planning Commission in favour of research, but, within a few days of the announcement of the award, he was named as the Prime Minister's science adviser and apparently decided to remain on the Commission as well.

According to an INSA spokesman, Menon's letter accepting the award, originally carrying a salary of 8,000 rupees a month, a contingency fund and the services of a research assistant, was received only in August this year, after reminders from academy officials, including Dr A.S. Paintal, the president since January 1987.

Menon will keep the Raman professorship until 1991, drawing a salary of one rupee a year. K.S. Jayaraman

* Shawyer, Colin. *The Barn Owl in the British Isles — its Past, Present and Future* (The Hawk Trust, Freepost, Beckenham, Kent, £6.95.)

City not to blame for poor science investment by British industry

London

THE failure of British industry to improve on its poor record of investment in research and development cannot be blamed on shareholders anxious to see high short-term profits at the expense of long-term stability, according to a survey published last week. It is often the companies themselves that neglect to give sufficient weight to long-term development, mainly because of underlying economic and political factors, including inadequate profitability.

The popular argument is that stock markets lay too much weight on current profits and current dividends. Companies respond to this pressure by pursuing short-term profit goals at the expense of longer-term investment.

After a year-long study, a 29-man task

from current investment.

A poll among company chief executives showed that financial market activities were not the most important constraining influence on companies' investment plans. Most felt constrained by either the cost of capital or fear of an inadequate return on their investment.

Another common perception is that the City is not prepared to raise funds for small businesses or high-risk ventures. Not so, say Nickson and his team. Some £396 million of equity and loans were provided by venture capital companies to UK companies in 1986, in amounts of between £100,000 and £1 million, compared with £279 million in the previous year and £110 million in 1982. All the main banks and financial institutions have their own venture capital funds.

What the system needs, the report concludes, is improved communications.

Companies must make more effort to keep the market informed. Shareholders and analysts must be given more information about companies' innovation activities. The government has indicated that it would like to see companies routinely reporting research and development expenditure in their annual accounts. Although the government says it would prefer a voluntary system, it does not exclude the possibility of legislation.

The CBI task force is opposed to statutory disclosure, and notes that for single-product companies and those that do not carry out research and development regularly, disclosure could mean revealing commercially sensitive information. The task force would support the introduction of a statement of standard accounting practice which leaves open the possibility of not disclosing if there is a sound reason. The disclosed figure of research and development spending should be put in its proper context by explaining how the research and development fits in with a company's overall strategy.

Simon Hadlington



Sir David Nickson — debunking mythology.

force of the Confederation of British Industry (CBI), chaired by CBI president Sir David Nickson, found that many of the commonly held perceptions surrounding the relationship between industry and the City were "simply not supported by the available facts". They were part of a "pervasive mythology that needs to be debunked in the interests of both the City and industry alike". The task force found that most financial institutions (such as pension funds) are long-term investors — "not surprisingly in view of the long-term nature of their obligations".

At the heart of the problem, according to the task force, lies a communications gap between management, owners of companies and those providing finance. British industry invests less than one per cent of the nation's gross domestic product in research and development, less than its major competitors. But no evidence exists to attribute this to 'short-termism', the report says. Indeed, industries where long-term investment is crucial, such as oil and pharmaceuticals, enjoy high price earning ratings, reflecting the long-term growth prospects expected

California gives radiation-safety violations a short half-life

San Francisco

CALIFORNIA researchers are suddenly acutely aware that they are personally liable to prosecution for radiation violations by members of their laboratories. This lesson has sunk in after criminal charges were brought against 10 members of the University of Southern California (USC) for violations of the radiation regulations. Although the charges against individual investigators have been dropped, the effects of the case continue to reverberate through the scientific community.

The USC case began in 1986, when a routine state radiation inspection found violations in several biomedical laboratories. Most were minor and, according to the university, posed no health threat. But the state turned its investigation over to the Los Angeles city attorney, who last March filed 179 criminal misdemeanour charges against the university and individual investigators. One investigator was charged on 21 counts, each carrying a maximum penalty of a \$1,000 fine and/or six months in gaol.

Steve Auer, attorney for USC, said he was surprised that the city attorney had filed criminal charges against the individuals. He had expected civil charges, which carry lesser penalties and less stigma, to be filed against the university. University of California (UC) attorney, John Lundberg, agreed, adding that to charge individual investigators is "grossly unfair". Lundberg said there was never any talk of charges against individuals

when state inspectors found 33 radiation violations at UC San Francisco last March. Civil charges against UC have been threatened but not filed in that case, he said.

But the Los Angeles district attorney's office stands by its action. "The university as a corporation is a fictional entity", said spokesman Steve Tekosky, "and it can only violate the laws as individuals violate the laws." Tekosky said he knew of no other case in which individual university faculty members have been held responsible for radiation violations.

On 1 October, all but 15 of the charges were dropped, including all of those against individuals. USC pleaded no contest to the remaining charges and paid a \$25,000 fine. But the state's initial harsh action achieved its goal: USC overhauled its radiation safety programme and has satisfied the city attorney's office that it is now in total compliance with regulations.

Researchers around the state are expressing alarm at the criminal indictments at USC by signing a petition denouncing the use of such "exaggerated measures". Copies have begun to arrive at the office of the state legislature's committee on science and technology.

The petition supports the need for effective enforcement of safety regulations, but criticizes criminal indictments as "regulatory over-reaction". It also notes that such draconian measures may force researchers to consider positions in "more enlightened" states.

Marcia Barinaga

But do they learn anything?

SIR—Michael Shortland (*Nature* 328, 213; 1987) used as part of his argument about the limitations of interactive science centres a quotation attributed to me in a daily newspaper. He then proceeded to question in a facile manner the validity of what Technquest, and indeed other interactive centres, are trying to do. I am well aware of the difficulties to which he draws attention, but a balanced discussion of these difficulties is not necessarily the main interest of a journalist in the limited space and context of a short article in a daily newspaper.

I therefore offer a longer quotation from the "Director of Technquest" (myself):

So visitors to interactive science and technology centres enjoy themselves. But do they learn anything? We all have a fund of anecdotes to illustrate that learning can occur, and many of us have the faith that the fun and enjoyment could well be a vehicle for alleviating the alienation from science felt by many people . . .

. . . [However] we have not yet invented a sufficiently diverse menu of programmes, methods and devices which can be incorporated into a coherent educational strategy whereby visitors would be allowed, enabled and encouraged to learn more if they so choose.

The task is to create the dishes on this menu and to implement educational strategies based on them. Only then will interactive exhibits deserve to be called 'plores' and interactive science and technology centres be able to fulfill their potential as part of the non-formal education scene.

JOHN BEETLESTONE

Department of Education,
University College,
Cardiff CF2 4AG, UK

SIR—Michael Shortland's comments (*Nature* 328, 213; 1987) on hands-on learning with interactive demonstrations and experiments, as in the Exploratory in Bristol, point to potentially exciting educational research. What indeed do visitors actually learn hands-on? Can ordinary people in such an environment discover truths of science for themselves? Can effective learning be fun, even to playing games? Perhaps we will find some answers in the Exploratory.

It is delightful watching families playing with the 'plores', for exploring, and questioning and explaining phenomena to each other. Their behaviour sometimes looks random, sometimes directed, but seldom totally inappropriate or destructive. It is hard to believe that learning has to be a solemn business, though quiet places for thought are essential.

Bertrand Russell's distinction between knowledge by acquaintance and knowledge by description sets up the significance, and also difficulties of presenting science

hands-on. Experiencing the forces of a large gyroscope is 'acquaintance' more evocative than any words; but it needs concepts and words to explain, for example, the precision of a gyroscope (Newton's first law, point masses) or to explain how images are formed, or why mirrors reflect. It is not clear how as children we move from experience to understanding, with concepts and propositions of language. Here the difficulties for any science museum start, for children, and most adults, simply do not have the necessary background concepts for explanations to be meaningful. We attribute this to the lack of exploratories over the past many years.

Successful exhibits or plores evoke some surprise. Surprising experience carries information, and suggests questions. Questioning, in turn, leads to seeking explanations and concepts which, especially when shared, may be expressed in words. Without shared experience and some understanding, the explanations — which through brevity lack the lead-uns and helpful hints of books — may be totally meaningless, however well written. This is like the neurological condition of agnosia ('lack of knowledge') which renders perception meaningless even though the eyes are not optically blind. The significance of phenomena and principles of how things work are simply invisible without necessary knowledge. The aim of exploratories is to cure the agnosia of virtually the entire population, by offering first-hand acquaintance of phenomena, and principles of science and technology, mainly through simple experiments and games. It is indeed remarkable that people willingly pay to play with pendulums, to discover underlying principles such as inertia and gravity. The hope is that through this interactive experience, many people will lose their agnosia blindness as they gain the knowledge and concepts necessary for seeing as science understands.

It turns out that plores on how we see and understand — including illusions of various kinds — can be particularly effective as they bridge experience and physics. We may hope that hands-on exploring leads to using the words and special symbols of science fluently, as acquaintance with phenomena grows to friendship with the world out there waiting to be discovered.

RICHARD GREGORY
(Hon. Science Director)

The Exploratory, Queen's Road,
Victoria Rooms, Clifton, Bristol, UK

Animal models

SIR—In the News and Views item 'Ironies of animal modelling', J.D. Hallonquist and N. Mrosovsky follow carefully the

psychiatric tradition of *post hoc* reasoning in their attempts to justify the use of animal models in the study of seasonal depression (*Nature* 329, 18; 1987). After admitting to the physiological inadequacy of the animal model, that is, ". . . it seems likely that the mechanisms for these [human] therapeutic effects are very different from those operating in the photo-periodic animal models . . .", and the very real possibility of the placebo effect, the authors simply accept the efficacy of the treatment. "Now", they seem to reason, "given the veridicality of the treatment, the use of the animal models must have been responsible". I can only hope this rebuttal reaches them before the advent of winter destroys the final threads of their precarious logic.

To my knowledge, the Hawthorne study remains the classic example of the placebo effect. If Hallonquist and Mrosovsky insist on argument by analogy, I suggest they re-read this study (under very bright light, of course) and leave the human-animal analogue to the equally implausible but more phenotypically congruent hypotheses of the physiologist.

DONALD J. BARNES

The National Anti-Vivisection Society,
112 North Carolina Avenue, SE,
Washington, DC 20003, USA

Conservation laws

SIR—We who are, so to speak, your lay readers, must thank you for reporting (*Nature* 328, 755; 1987) the flurry among the high priests of physics on how to interpret the conservation laws for mass and energy. You seem to suggest that the correspondence in *Physics Bulletin* "has been closed" because it will not do for the acolytes, not to mention the general public, to hear the squabbling.

The need for care in the formulation of the conservation laws is well expressed in R. O. Kapp's last book *Towards a Unified Cosmology* (Hutchinson, 1960). Kapp was a distinguished engineer and teacher, well acquainted with relativity theory. He offers two statements for the principle of conservation of energy, the second of which reads: "The total energy in a self-contained system is constant."

He is, however, at pains to distinguish between a "self-contained" and an "isolated" system. In Kapp's view, the material Universe should not be regarded as a self-contained system. "This is just one of the many sly hypotheses that cause misconceptions. They enter our deliberations so unobtrusively that their entry passes unobserved. Thereupon they clothe themselves in the garments of an irrefutable fact. The disguise is often so excellent that even the most critical do not penetrate it."

D. H. TOMPSETT
34 Vicarage Road, Henley on Thames,
Oxfordshire RG9 1HW, UK

Shapes of quasi-crystallites

Although the sensation has evaporated from the search for more exotic crystals with five-fold symmetry, there may yet be unexpected phenomena to be described.

WHATEVER can have happened to quasi-crystals, the curious metallic compounds of aluminium and manganese that startled us all two-and-a-half years ago by revealing that five-fold symmetry can, indeed, arise naturally? That the issue of what these materials amount to microscopically has run into the sand surprises nobody. Many more examples of the phenomenon have been found, in ternary alloys such as U-Pd-Si, for example.

So it is strange that nobody can as yet put his hand on his heart and say that the explanation is a simple case of a three-dimensional generalization of a two-dimensional Penrose tiling of the plane (when two rhombi provide close cover), whether a different mechanism such as Pauling's twinning is as good an explanation of how the whole of three-dimensional space might be filled with atoms (given the possibility of some distortion) or whether the truth is that quasi-crystals are microscopic regions of ordered lattice with five-fold symmetry linked together with amorphous regions. No doubt the question of what is what will eventually be sorted out.

Almost as in despair, attention has therefore turned to the task of telling whether something can be learned about the internal arrangements of atoms from the external shapes of the crystallites their assemblies form. The technique is nicely illustrated by the case of rock salt (NaCl), which will form cubic crystallites, or regular octahedra, or almost any shape derived from the principle that a surface facet will be densely packed with atoms of some kind. The shapes of the facets can be classified by the rotational symmetry groups ('point groups' as they are called) which are subgroups of the overall symmetry of the crystal.

Icosahedral crystallites behave as might be expected. Just as the X-ray diffraction patterns of a quasi-crystalline lattice show physically impossible five-fold symmetry, so the shapes of the microfacets do not belong to physically realizable point-groups. This seems to show that the underlying ordered condition of the atomic lattice is similarly one that should not be allowed in a strictly periodic lattice which is infinite in extent. The key observations seem to be those of Dubost, Lang, Tanaka, Sainfort and Audier (*Nature* 324, 48; 1986).

This is where the theoreticians enter. If the allowable aggregates of atoms in a

quasi-crystallite are relatively small, should it not be possible to calculate from first principles the shapes of the aggregates they are likely to assume, and thence to say something of the underlying arrangement of the atoms. By coincidence, two very similar but clearly independent calculations of the shapes of micro-quasi-crystals appear in the current issue of *Physical Review Letters* (59, pp. 1679 and 1683; 1987). The essential conclusion appears to be that crystals do indeed have faceted surfaces when the temperature is that of absolute zero, but that there is a "transition to roughness" at some higher temperature.

Part of the interest of these publications is that they demonstrate that Dov Levine from the Institute of Theoretical Physics at the University of California, Santa Barbara, and one of the veterans of the quasi-crystalline business, has stayed with the problem. The simplest case is that in two dimensions, where the Penrose tiling is a sufficient representation of a two-dimensional crystal. The Penrose tiling is that which shows that the plane can indeed be covered exactly by a suitable arrangement of only two rhombi whose diagonals stand in the ratio of the golden mean (half of 1 plus the square root of 5) or of its reciprocal.

The difficulty in making any calculations of quasi-crystals is that of finding a suitable notation. Levine and his colleague Anupam Garg (*Phys. Rev. Lett.* 59, 1683; 1987) follow Levine's original technique of using a set of five unit vectors in the plane, each of them figuratively directed towards the vertices of a regular pentagon.

To calculate something about the energy of a quasi-crystallite, it is only necessary to observe that there are, indeed, imperfect (zig-zag) cleavages that may be drawn through a two-dimensional Penrose lattice simply because the possible orientation of the edges of the tiles are finite in number, and because there is also a degree of translational symmetry in the problem.

By supposing that each tile is occupied by an atom, which interacts with its neighbour by means of a fixed energy, Gard and Levine show that the least energy for a planar crystalline has the shape of a regular decagon. In three dimensions, building blocks are two differently shaped rhombohedra, and the crystallite with the least energy is

the rhombicosidodecahedron (a dodecahedron with decorated surfaces).

Christopher L. Henley (Cornell University) and Reinhard Lipowsky (Nuclear Physics Institute, Jülich) come to essentially the same conclusion in the other paper (*Phys. Rev. Lett.* 59, 1679; 1987), but by a different argument which, interestingly, suggests links between the calculation of the shapes of crystallites with the least energy and the band structure of electrons moving in quasi-periodic lattices.

Nobody will scoff at the problems of carrying through these cumbersome calculations, which entail the awkward circumstance that it is necessary repeatedly to check the conclusions of the algebra against intuitive impressions of how two-dimensional tilings appear. One startling conclusion, however, is that the surface energy of a cleavage plane drawn through a quasi-lattice may actually be less than that expected for a regular lattice of similar density. Is that, people will be asking, why it is very small crystallites commonly appear with quasi-crystalline lattices?

The notion that there may also be a transition to roughness as the temperature is increased is the chief surprise of these calculations. Both sets of authors are eager to point out that, so far as their algebra and numerical calculations go, even for the two-dimensional case, the results are valid only for temperatures which are very small. But in the case of three-dimensional solids, the transition temperature could easily be finite, implying that regular micro-quasi-crystallites would be observed only under the extreme conditions that they are formed at absolute zero.

The physical consequences of these predictions are not immediately apparent. Might the calculations help, for example, to account for the true structure of three-dimensional crystallites as roughened crystallites held more or less mechanically together? Certainly, according to the calculations in these two papers, there should be circumstances in which internal fragmentation happens almost as a matter of routine.

But the chief interest of these articles, so far as they go, is that they show how much life still remains in the problem of quasi-crystals. Many people, no doubt, will be grateful for that proof.

John Maddox

Antipeptide antibodies

A vaccine for the common cold?

Ann Palmenberg

ALMOST everyone has suffered, at one time or another, from the sneezing, sniffling, watery-eyed misery of a good old-fashioned common cold. The culpable agent in many cases is a rhinovirus, one of the small, positive-sense RNA-containing picornaviruses, whose very name evokes images of a runny nose. Elsewhere in this issue (*Nature* 329, 736–738; 1987), Joseph McCray and Gudrun Werner report their recent work on producing a vaccine against one of these rhinoviruses.

Rhinovirus infections are among the most prevalent and acute of human upper respiratory illnesses. Annually, the disease exacts high tolls as cold-related absenteeism results in millions of lost work days. Advertisers constantly (and profitably) harp on the therapeutic value of antihistamines, throat lozenges, non-irritating tissues and even chicken soup to ease the symptoms and make life bearable until the cold goes away. But, let's face it, what the world really needs is not a better variant of aspirin. It needs a safe, effective prophylactic vaccine to prevent rhinovirus infections in the first place.

This idea is not new. Vaccines for other picornaviruses, such as polio and foot-and-mouth (FMD) disease, have been

around for many years. Unlike poliovirus, however, attenuated strains of rhinovirus have never been isolated, an experimental catch that precludes development of analogous Sabin-type vaccines. The present vaccines for FMD, with more limited effectiveness, rely on inactivated virus, prepared separately for each of the seven viral serotypes. This approach, too, is inappropriate for rhinoviruses because of wide serotypic diversity. A recent collaborative report on rhinoviruses (Hamparian, V.V. *et al. Virology* 159, 191–192; 1987) disclosed that at least 100 distinct serotypes of human rhinovirus have already been characterized. Even if a safe, potent, inactivated rhinovirus inoculum could be prepared, the sheer number and distribution of serotypes create enough logistical and licensing problems to dissuade any pharmaceutical firm. Clearly, a different approach to rhino vaccines is necessary.

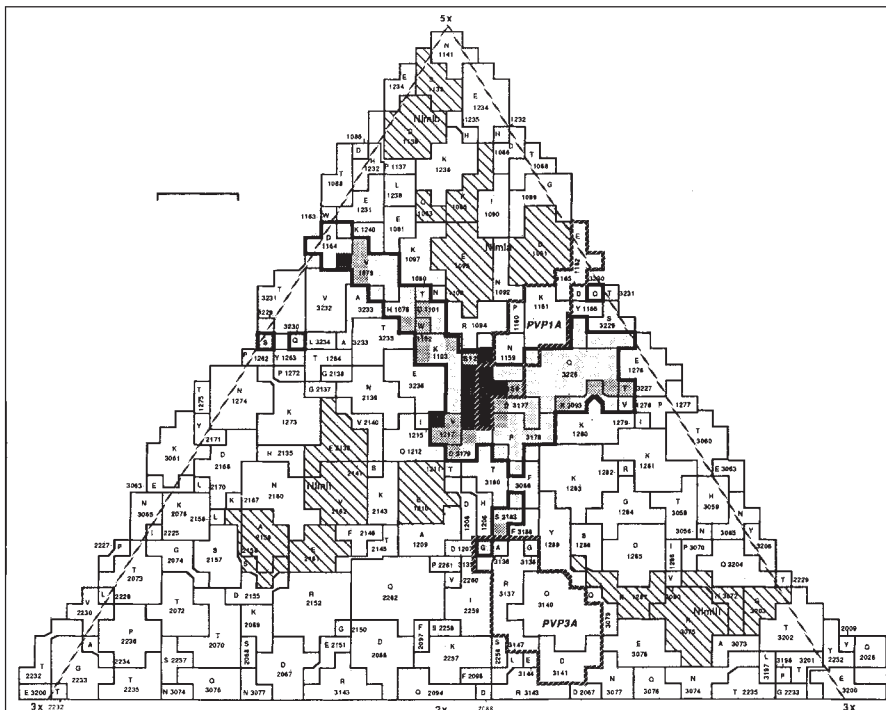
The new work by McCray and Werner reported in this issue describes preliminary findings with antipeptide antisera developed against human rhinovirus type 14 (HRV14). The authors chose this particular virus because the three-dimensional crystal structure of HRV14 has been determined to atomic resolution

(Rossmann, M.G. *et al. Nature* 317, 145–153; 1985) and serves as a powerful predictive model for the amino-acid residues on the surface of the virus that have antigenic potential. McCray and Werner prepared two synthetic peptides corresponding to amino acids 147–162 in the virion capsid protein VP1 (peptide PVP1A) sequence and to residues 126–141 of virion capsid protein VP3 (peptide PVP3A). From Rossmann's structural data, they reasoned that these sequences might form part of a hypothetical cellular receptor binding site, or 'canyon', on the particle surface. Therefore, antibodies against the peptides should prevent receptor binding and effectively neutralize the virus.

Indeed, polyclonal sera raised in rabbits against the synthetic peptides not only bind HRV14 particles in enzyme-linked immunosorbent assays (ELISAs), but at appropriate dilutions, actually neutralize virus. More important, the antisera also have activity against 27 other rhinoviruses, representing about 60 per cent of the tested serotypes. The antisera are not effective against polio, echo 9, coxsackie B3 or rhinoviruses from the minor receptor group (a distinct receptor binding group comprising about 20 per cent of all rhino isolates). Further, McCray and Werner show that those antibody fractions with demonstrable binding activity against both synthetic peptides also have enhanced titres against HRV14. They propose that a short amino-acid sequence (PPGA), common to both peptides and possibly conserved among other rhinos, could be responsible for eliciting these more effective neutralizing antibody populations. They optimistically hope their results "... suggest the feasibility of a synthetic peptide vaccine for rhinoviruses ...".

It sounds great, doesn't it? But don't throw away your chicken soup just yet. The neutralizing titres of the rhino antipeptide sera are at least four orders of magnitude lower than typical hyper-immune sera produced by inoculation with intact HRV14. For many of the 27 other 'positive' serotypes, the titres are even lower. Over the past few years, virtually identical peptide experiments have been carried out with FMD, polio and hepatitis A viruses. Invariably, the elicited antibodies are positive for binding in ELISA tests, but have very low neutralizing titres, and ultimately confer little or no protection against virus challenge to inoculated animals.

The most promising FMD experiments (DiMarchi, R. *et al. Science* 232, 639–641; 1986) used selected peptides 40 amino acids long to produce neutralizing titres about 1,000 times higher than those of McCray and Werner. Although some cattle were protected from virus challenge by peptide inoculation, the circulating levels of peptide-induced antibodies could not be correlated with the specific degree



Amino-acid residues on the surface of an HRV14 icosahedral asymmetric unit are displayed schematically in projection on a grid at 2-Å intervals. The grooved canyon is highlighted, and residues lining its sides are shaded according to depth (darker = deeper). The hatched residues mark the positions of immunogenic neutralization escape mutants (NIMs), according to the work of Rossmann *et al.* Exposed residues within peptides PVP1A and PVP3A are indicated. Scale bar, 10 Å. (Adapted from a manuscript in preparation by M.G. Rossmann and myself.)

of protection in individual cattle in the study of DiMarchi *et al.* Conventional FMD vaccines, in contrast, have more predictable responses. They tend to elicit a slightly lower neutralizing titre than the best synthetic peptides, although all this titre is apparently protective. Based on animal tests, many researchers have limited enthusiasm for the prospects of effective picornaviral peptide vaccines.

Other specific problems with this approach for rhino vaccines lie within the structure of the virus itself and with the distribution of its natural epitopes. The figure schematically shows the surface residues of an HRV14 crystallographic subunit. The putative receptor-binding canyon is far too narrow to permit entry by an antibody, whose arms average 20–30 Å in diameter. Amino acids lining the edge of this groove are somewhat more accessible, although they are also more variable among the serotypes. On virions, only limited portions of six residues from each of PVP1A and PVP3A are exposed to the surface. The common PPGA fragment of VP3 is deeply buried, and in VP1 the equivalent segment lies near the bottom of the canyon. It is therefore unlikely that this particular sequence can be the primary determinant of neutralization. Rather, the rhino anti-peptide neutralization activity is probably caused by limited steric interference with receptor binding as antibodies react with the 4–7 exposed amino acids near, but not completely in, the canyon. Because this type of interaction, predictably, should be weak, the low neutralizing titres of these peptide-induced sera are not surprising.

The sequences of PVP1A and PVP3A are different from all the mapped, natural immunogenic residues of HRV14, which are located on the more exterior portions of the particle surface. But would peptides containing these regions be any more effective? Probably not. The neutralizing immunogenic sites of this virus, and probably of most other picornaviruses, are generally made up of discontinuous sequence fragments brought together on the surface by the overall structure of the virion. A single amino-acid mutation can alter the neutralization characteristics of a particular isolate (see Rossmann *et al.*). Even if our synthesizers could keep up with evasive changes by the viruses, peptides that mimic both the relevant sequence and structure of the rhino epitopes are beyond our present laboratory expertise.

Sadly, this means that until someone has a better idea, the best advice when you have the sniffles is to stay in bed, drink plenty of liquids and call your doctor in the morning. It will be some time before we can cure the common cold. □

Ann Palmenberg is in the Institute for Molecular Virology and Department of Veterinary Science, University of Wisconsin, Madison, Wisconsin 53705, USA.

Avian ecology

Adaptive variation in bill size of African seed-crackers

Peter Boag

DESPITE the scarcity of convincing examples, the idea that variation in the foods available to birds can promote intraspecific variation in trophic morphology, such as bill size or tarsus length¹, remains attractive, and theories about the evolution of such phenomena flourish². The debate will be revitalized by Thomas Bates Smith's field study of *Pyrenestes* finches in equatorial Africa, reported on page 717 of this issue³. Smith finds that three congeneric species of *Pyrenestes* each display pronounced bimodality of bill size, independent of sex, age, body size or geographical origin. The variation is correlated with differences in feeding efficiency on hard and soft seeds. There is no evidence for assortative mating by bill size, but preliminary data suggest that the variation has a genetic component.

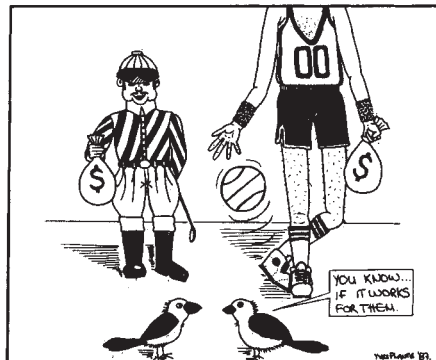
Van Valen¹ was one of the first champions of the 'niche-variation' hypothesis. This theory contrasts the traditional view that populations contain a single optimum phenotype conferring maximal fitness, with the alternative that populations contain polytypic individuals with equal fitnesses in different parts of the available niche. Early tests of the hypothesis compared morphological variation of birds occupying wide versus narrow food niches (such as island and mainland populations, respectively). The prediction that the absence of competitors on islands would select for individuals with a greater range of trophic morphologies has met with mixed results when tested². The high mobility of most birds results in relatively large effective population sizes, with little genetic differentiation between local populations. Also, there are few examples of assortative mating based on trophic morphology. Results such as these make it unlikely that genetic polymorphisms in trophic morphology are maintained by disruptive selection alone.

Despite these concerns, the idea that intraspecific morphological variation is adaptive remains intuitively attractive. Analogies with our own species spring to mind, if we draw parallels between economic success and 'fitness'. Although they possess atypical phenotypes, a diminutive jockey and a towering basketball player can each command a huge annual income, far greater than the population mean. It is the very rarity of their unique physical attributes that allows such individuals to exploit a specialized professional 'niche'.

Although the extremely variable

Pyrenestes species have baffled taxonomists for years, Smith³ is the first to examine museum specimens in conjunction with a field study of one species (*P. ostrinus*). His field data from Cameroon clearly show two nonoverlapping forms, each showing considerable variation. The museum data for other populations are much fewer. In most cases it is difficult to identify a discrete group of larger individuals, but all three species of *Pyrenestes* that Smith examined span a remarkably wide range of bill sizes.

Smith argues that the functional significance of the bill-size variation is in the differential use of niches. Small-billed birds prefer and are most efficient at eating small seeds; larger individuals, larger seeds. This conclusion agrees with findings for Darwin's finches, as well as for other birds⁴. But it remains unclear whether disruptive selection can be strong enough to maintain discrete bill morphs



without assortative mating or other constraints on gene flow.

Although the empirical demonstration of this pattern is intriguing, it is more difficult to interpret its evolutionary significance. Smith suggests the variation has a genetic basis, and although he does not present an explicit genetic model, he interprets the large and small progeny produced by a single mixed pair in captivity as supporting the existence of a genetic polymorphism. This conclusion seems to assume mendelian inheritance (for example, a single locus, two-allele system with dominance, and at least one heterozygous parent). But virtually every study to date that has looked at the genetics of beak size supports a polygenic model². One of the rare instances in which discrete genetic variation has been associated with size variation is the unusual chromosomal polymorphism found in some North American sparrows (*Junco* and *Zonotrichia*), where karyotype is correlated

with size, plumage and behaviour⁵. Also, the fact that in *Pyrenestes* the bill size varies so much more than other morphological characters indicates strong allometry. In most birds, external morphological characters covary positively, so that very strong, antagonistic selection is needed to produce large changes in one character alone⁶.

It is possible that the polymorphism is a threshold trait, with polygenic inheritance, or that the morphs do not differ genetically at all. For example, they could represent alternative developmental pathways⁷, triggered by environmental cues such as which rainy season they were reared in, hatching order or food availability. The morphological polymorphisms found in insects stem from various mechanisms, ranging from purely environmental determination⁸ to single-locus and polygenic inheritance⁹. Smith's paper in this issue focuses on a mendelian explanation for the morphs, implying that genetic differences between individuals are needed for the evolution of adaptive variation. In many situations, selection may instead favour a system of genetically

similar individuals that express environmentally sensitive, plastic phenotypes. Similar complications arise in discussions of the evolution of alternative reproductive strategies, as pointed out by Austad⁹.

These issues can be settled only by additional breeding experiments, probably on captive animals, as the study of wild *Pyrenestes* broods is difficult. Such experiments are the only way to decipher the mechanism behind this remarkable variation in seed-cracker bills. Once the mechanism is understood, whether or not it is a mendelian polymorphism, the challenge will be to explain how the variation originated and is maintained in this genus. □

1. Van Valen, L. *Am. Nat.* **99**, 377–390 (1965).
2. Grant, P.R. & Price, T.D. *Am. Zool.* **21**, 795–811 (1981).
3. Smith, T.B. *Nature* **329**, 717–719 (1987).
4. Grant, P.R. *Ecology and Evolution of Darwin's Finches* (Princeton University Press, 1986).
5. Rising, J.D. & Shields, G.F. *Evolution* **34**, 654–662 (1980).
6. Smith-Gill, S.J. *Am. Zool.* **23**, 47–55 (1983).
7. Bernays, E.A. *Science* **231**, 495–497 (1986).
8. Roff, D.A. *Evolution* **40**, 1009–1020 (1986).
9. Austad, S.M. *Am. Zool.* **24**, 309–319 (1984).

Peter Boag is in the Biology Department, Queen's University, Kingston, Ontario, Canada K7L 3N6.

Mathematics

The arithmetic of chaos

Ian Stewart

"MATHEMATICS", said Carl Friedrich Gauss, "is the queen of the sciences, and arithmetic is the queen of mathematics." By arithmetic he meant the theory of numbers, not $2+2=4$, and the tendency of queens not to dirty their lilywhite hands was not entirely absent from his mind. The overt subject matter of number theory — the patterns and perplexities of ordinary whole numbers — does not evoke immediate applications to science. "That subject is in itself one of peculiar interest and elegance, but its conclusions have little practical importance", wrote W. W. Rouse Ball in 1896. In terms of the common division of mathematics into 'pure' and 'applied', number theory is about as pure as you can get: poles apart from traditional applied topics such as dynamics. Not any more.

For at least a decade, evidence of the applicability of number theory has been accumulating. Manfred Schroeder's *Number Theory in Science and Technology* (Springer, Heidelberg, 1985) makes an eloquent case. And now Ian Percival and Franco Vivaldi of Queen Mary College, London (*Physica* **25D**, 105–130; 1987) have discovered extensive applications of deep and abstract concepts of number theory to one of the most flourishing, fascinating and fashionable research fields of mathematical physics — chaotic dynamics. (See the News and

Views article by Tomas Bohr and Predrag Cvitanović, *Nature* **329**, 391–392; 1987.)

Their methods are taken from algebraic number theory — meaning the theory of algebraic numbers rather than the algebraic theory of numbers. Let me describe this first. A complex number is algebraic if it satisfies a polynomial equation whose coefficients are rational numbers; it is an algebraic integer if those coefficients are whole numbers. For example, the golden number $\varphi = (1 + \sqrt{5})/2$ satisfies the equation $\varphi^2 - \varphi - 1 = 0$.

Associated with φ is its number field, consisting of all numbers $p + q\varphi$ where p and q are rational, and its ring of integers, for which p and q must be whole numbers. Algebraic number theory studies the arithmetic of these rings of algebraic integers. The central concept is that of an 'ideal', an abstract gadget introduced to get round the awkward tendency of algebraic numbers to go misbehave wildly when factorized into primes. It originated in the work of Ernst Eduard Kummer aimed in part at solving Fermat's last theorem: the equation $x^n + y^n = z^n$ has no integer solutions when $n \geq 3$. The theorem remains unproved to this day, although recent work suggests that it may not remain unproved much longer.

So much for the techniques; now for the subject matter. Chaotic dynamics is a dramatic and exciting new field on the

boundary between mathematics and the physical and life sciences. It begins from the belated observation that purely deterministic systems can behave in random ways. Chaos challenges the entire philosophy of classical mathematical modelling.

A major area in which chaos plays a crucial role is hamiltonian systems, the study of dynamics in the absence of friction. The modern approach to hamiltonian systems is geometric. The position and momentum variables of the system define a multidimensional 'phase space'. A bicycle, for instance, can move with (at least) five degrees of freedom. The front wheel can revolve and point in any direction, while the back wheel and gear system, and the two pedals, can revolve independently. The position of the bicycle corresponds to a point on a five-dimensional torus, and the various angular momenta live in a five-dimensional euclidean space. The phase space of a bicycle combines these two, and is thus ten-dimensional. The motion of the bicycle is determined by a single representative point, which wanders through this ten-dimensional phase space, obeying newtonian dynamics. By considering all possible motions of the bicycle, we define such a path for each initial point in the phase space; these paths are analogous to those traced by particles suspended in a fluid. Thus dynamics is envisaged as the flow of an idealized fluid through a multidimensional space.

Classical applied mathematics describes this flow when its motion is sufficiently regular. Henri Poincaré was the first to discover that your typical hamiltonian-system-in-the-street can behave in a very irregular manner indeed. Horrified by what he saw, he took the analysis no further. The Russian mathematician Vladimir Arnold has rightly described the analysis of a typical hamiltonian system as "beyond the capability of modern science." The reason is that in the phase space of such a system there are 'elliptic islands' of regular behaviour and 'homoclinic tangles' of random behaviour, mixed together like meatballs and spaghetti in an almost indescribably complex fashion — dynamics bolognaise.

Percival and Vivaldi start from the observation that "two special types of limiting system are free from this pathological structure, namely the integrable systems which are purely regular, and the Anosov systems which are purely chaotic". They confine their attention to the latter, concentrating on the comparatively neglected but important case of discrete dynamics.

A discrete dynamical system starts in some initial configuration and evolves at fixed instants 1, 2, 3, ... of time according to some fixed rule. The simplest behaviour occurs when the initial configuration remains unchanged, in which case it is called a fixed point or steady state of the

dynamics. The next simplest behaviour is a periodic cycle of states, repeating some finite series of configurations without end. Beyond this comes quasiperiodic behaviour, where several independent periodicities are in effect superimposed: the system almost repeats its evolution.

Chaos is dramatically different: it is effectively unpredictable and appears to be totally random. But even in a chaotic system, there are some initial states that evolve with a precise periodicity. They elude observation because they are unstable, and surrounded by states that evolve irregularly; yet they form a regular skeleton about which the flesh of chaos hangs. As such, they make some of the chaos vulnerable to attack by a determined mathematician.

Percival and Vivaldi aim to understand the arithmetical properties of the periodic states of a purely chaotic system. Do states of a given period exist? How many? Where? As phase space they take a two-dimensional torus — in classical terms, one position and one momentum coordinate — and the dynamics is obtained by iterating a mapping of this torus on to itself.

Let me give an example. A point on a torus is determined by its coordinates



(x, y) , where x and y lie between 0 and 1. The mapping that rejoices in the name of Arnold's cat sends each point (x, y) to $(x+y, x+2y)$, where these numbers are replaced by their fractional parts if they are larger than 1. Geometrically, the torus is stretched, rotated, and laid down on itself. (Arnold illustrated the geometry by drawing a cartoon cat on the torus, hence the name.)

The eigenvalues of this operation, which measure the amount of stretching that occurs along its principal directions of expansion or contraction, are algebraic numbers. Percival and Vivaldi show that many features of the dynamics of the mapping can be rephrased as questions about the arithmetic of the eigenvalues. For the cat map, the golden number φ is an eigenvalue: the dynamics of Arnold's cat is determined by the arithmetic of the golden number.

The methods work on any generalized cat map with integer coefficients, defined by a 2×2 matrix that satisfies a technical condition called hyperbolicity.

The nature and number of periodic points of a generalized cat map can then be deduced — more efficiently than by brute computation — by applying large chunks of the classical machinery of algebraic number theory. These include ideal theory, ideal classes, the quadratic reciprocity law, norms, units and the classification of primes as ramified, split and inert. Percival and Vivaldi say in summary: "Most of the ideas we use here were developed before the end of the nineteenth century, yet they have had little impact on the development of modern nonlinear dynamics. These methods not only provided effective computational tools, but also shed some light on the intimate structure of irregular orbits, which share with the theory of algebraic integers the same richness and unpredictability."

Mathematics is full of surprises, as this newly discovered link between Fermat's last theorem and dynamics makes plain. What is particularly curious is that the connection is both natural and elegant. The flow of results is so compelling that it would be easy to underestimate the depth of insight that lies behind them. It is a marvellous example of what can be achieved by putting together known but apparently unrelated techniques. Some inspired delving in an unlikely corner of the mathematician's rag-bag, and another implement takes its place in the physicist's toolkit. Who knows what else may be in there, neglected in the dark, awaiting its moment of glory? □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Vascular physiology

A family of angiogenic peptides

Judah Folkman and Michael Klagsbrun

WHAT are the factors that regulate the growth of capillary blood vessels? Interest in such angiogenic factors arose from experimental studies suggesting that tumour growth is angiogenesis-dependent and that various non-neoplastic diseases such as diabetic blindness and rheumatoid arthritis might be ameliorated if pathological angiogenesis were inhibited. In the past three years, four structurally unrelated polypeptides have been shown to be angiogenic (see ref. 1 for a review): fibroblast growth factor (acidic and basic FGF), angiogenin, transforming growth factor- α (TGF- α) and TGF- β . We can now add to the family tumour necrosis factor- α (TNF- α), described in two recent papers by Frater-Schroder and her colleagues² and by Leibovich *et al.*³

In both these studies, recombinant TNF- α was shown to be a potent angiogenic stimulator *in vivo*. TNF- α is known to be secreted by macrophages when these cells are activated by certain large molecules such as bacterial endotoxin or fibrin products. The study by Leibovich *et al.* suggests that TNF- α is one of the main angiogenic molecules released from macrophages. Macrophages are attracted into wounds and are thought to be essential for normal healing, so TNF- α may be involved in the regeneration of capillaries in wound healing.

The new blood vessels in a healing wound are formed by a series of sequential steps similar to the development of new blood vessels in many different processes such as ovulation and tumour growth¹. The events that are thought to be essential include: local degradation of the basement membrane of the parent vessel, allowing protrusion of endothelial cells;

outward migration of endothelial cells in tandem to form a capillary sprout; proliferation of endothelial cells within the sprout; and the formation of a lumen with subsequent branching. An *in vitro* analogue of each step has been developed as a bioassay, and these bioassays, together with methods of inducing neovascularization in the chick embryo and rodent or rabbit cornea, led to the discovery of angiogenic activity for each of the polypeptides listed above.

Once the sequential elements of capillary developments were known, it was generally assumed that any putative angiogenic factor could stimulate most, if not all, of these steps *in vitro* and *in vivo*. Indeed, this is true for the FGFs, the first angiogenic polypeptides reported. Angiogenic molecules discovered later, however, induce neovascularization *in vivo* but do not affect some or any of the *in vitro* bioassays. For example, although FGF is an endothelial mitogen *in vitro*, angiogenin is not, and TGF- β actually inhibits endothelial proliferation. Both the new studies^{2,3} report that TNF- α inhibits endothelial cell proliferation *in vitro*, although the peptide is not cytotoxic to endothelial cells *per se*. Thus, the *in vitro* and *in vivo* activities of TGF- β and TNF- α seem to be paradoxical.

This apparent paradox can be resolved if there are multiple pathways for capillary formation. Neovascularization, for example, could be accomplished by a polypeptide such as FGF which can 'directly' stimulate capillary endothelial cells to undergo one or more of the key steps in capillary growth delineated above. Similarly, angiogenesis could be mediated through an intermediate cell,



A new capillary sprout growing through collagen of the rabbit cornea towards a tumour implant. Lower right, leading endothelial cell. A lumen is forming in the middle of this sprout. (Magnification, approximately $\times 1,800$). The formation of such a sprout may be stimulated by an angiogenic factor diffusing from an extravascular site, but a very different effect may result from the intravascular injection of the factor (from ref. 8).

for which the macrophage is a good candidate. The angiogenic activity of TGF- β might be macrophage-dependent. TGF- β is a powerful chemoattractant for macrophages and might activate them to release a 'direct' angiogenic factor such as TNF- α , which is itself chemotactic to endothelial cells¹. Macrophages release other angiogenic factors as well², particularly under hypoxic conditions.

It is also possible that neovascularization is mediated by the release of angiogenic factors stored in tissues in an inactive form. Thus, the heparin-binding growth factor basic FGF is found in extracellular matrices of vascular endothelial cells as well as in basement membranes of cornea in association with heparin-like molecules³. FGF released by dissolution of extracellular matrix or by exposure of the extracellular matrix to heparin, possibly from mast cells, could initiate neovascularization.

Finally, the ultimate effects of an angiogenic polypeptide could depend greatly on its route of administration. Vascular endothelial cells are bipolar, with a luminal side and a basement-membrane side. When TNF- α is implanted in an extravascular location (basement-membrane side) capillary sprouts grow out of from existing vessels (refs 2, 3; see figure). In contrast, the intravascular injection of TNF- α (luminal side) may well stimulate procoagulant activity by endothelial cells as it does *in vitro*^{5,6}, but not capillary growth. Meth A sarcoma, for example, regresses when the host animal is injected with TNF- α (see ref. 7). There is a dramatic response in the tumour vasculature involving thrombosis, disruption of small vessels and haemorrhage into the tumour. This early event is probably the

procoagulant effect of TNF- α acting on the luminal surface of proliferating endothelium. This difference between the extravascular and intravascular actions of TNF- α may help to resolve the question of how this molecule is both angiogenic and tumour necrosing. Other growth factors may also have a differential effect on vascular endothelium, depending on whether the factor is administered inside or outside the vascular system. FGF, for example, induces neovascularization when it is implanted at an extravascular site, but it may have quite different effects when administered into the circulation.

The addition of TNF- α to the list of angiogenic polypeptides raises the question of why there are multiple angiogenic factors. Are they all relevant *in vivo*? B.R. Zetter (personal communication) suggests that angiogenesis is so essential, for example in wound healing, that it requires a high level of biochemical redundancy. Now that many angiogenic molecules have been identified, it may be fruitful to study how the known angiogenic molecules regulate blood-vessel growth *in vivo*. □

1. Folkman, J. & Klagsbrun, M. *Science* **235**, 442–447 (1987).
2. Frater-Schroder, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5277–5281 (1987).
3. Leibovich, S.J. *et al. Nature* **329**, 630–632 (1987).
4. Banda, M.J., Knighton, D.R., Hunt, T.K. & Werb, Z. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7773–7777 (1982).
5. Bevilacqua, M.P. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4533–4537 (1986).
6. Nawroth, P.P. & Stern, D.M. *J. exp. Med.* **163**, 740–745 (1986).
7. Old, L.J. *Science* **230**, 630–632 (1985).
8. Folkman, J. in *Important Advances in Oncology 1985* (eds DeVita, V.T. Jr *et al.*) Ch. 4 (Lippincott, Philadelphia, 1985).

Judah Folkman is Andrus Professor of Pediatric Surgery and Professor of Anatomy, and Michael Klagsbrun is Associate Professor of Biological Chemistry at the Departments of Surgery, Anatomy and Biological Chemistry of the Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.



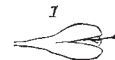
100 years ago

Fertilization of the Coffee Plant

YOUR readers are doubtless aware that coffee was cultivated some twelve years ago to a very large extent in Ceylon and South India, but owing to the attacks of leaf disease, the area has been rapidly reduced, except in some parts of Coorg and Mysore, where the climate is drier, and the leaves suffer less from the fungus. It has now been largely replaced by tea.

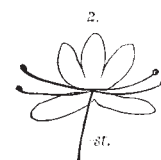
The jasmine-like flowers of the coffee are borne in clusters in the axils of the leaves, and appear simultaneously all over the estates. After a prolonged drought of one or two months, or even more, at the beginning of the year, there is generally a heavy fall of rain, sometimes lasting only an hour or two, sometimes continuing for two or three days: the amount that falls must be enough to saturate the ground, and not less than one inch.

In from six to eight days from the time of the first shower, the flowers burst into full blossom, last for a day, and then drop off. On the evening before the blossom is fully out, if the flowers are examined it will be found that they are partially

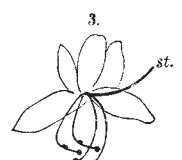


open, the stigma being protruded and receptive. During the night the hum of insects can be distinctly heard, and I am of opinion that the flowers are largely fertilized by night-flying insects which carry pollen from those flowers which happen to be open rather before the others, as some are delayed. On the following morning all the flowers will be found open, and the field of coffee presents a sheet of white. These flowers are frequented by immense numbers of bees, of two kinds, one about three-quarters of an inch long and black, the other smaller and with white bands round its abdomen. The stigmas now are covered with pollen, and the anthers bursting, and the larger of these bees may be seen buzzing from flower to flower sweeping up the grains of pollen between its front legs, and rolling them into balls. Long before evening all the anthers are exhausted of pollen, and the insects have departed. Besides bees some butterflies visit coffee.

While on this subject I may mention the curious alteration in the position of the organs of *Clerodendron infortunatum* when flowering. This plant is proterandrous: at first the style



First position.



Second position.

hangs down, while the stamens are erect: as soon as the pollen is shed the stamens drop, while the style rises, and the stigma becomes receptive. The chief carriers of pollen in this plant are small ants.

T. F. BOURDILLON.
From *Nature* **36**, 580; 20 October 1887.

Archaeology

Dark-age cities discovered

Richard Hodges

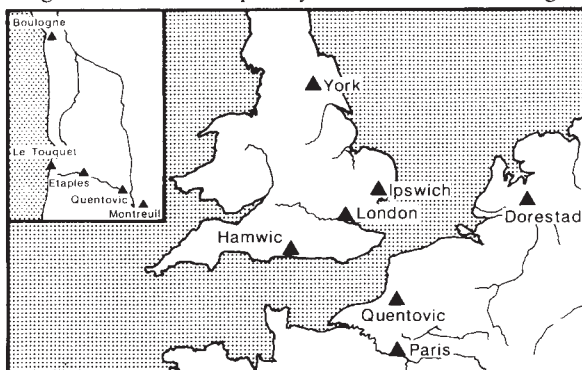
RECENT archaeological finds in several areas of Europe demonstrate the value of excavation rather than relying on contemporary written records for a complete understanding of mediaeval societies. Arguably the most important recent discovery¹ that provides a piece of the enigmatic jigsaw of our picture of seventh-century societies is that of Quentovic, a town near Montreuil in northern France.

The history of early mediaeval trade and commerce around the North Sea basin has been gradually pieced together from an unexpected series of discoveries. The great port of Dorestad, located in the Rhine estuary, was found by starving peasants rooting around the early mediaeval bone-pits (as they described them) and boiling up their contents. Subsequent excavations revealed that this town covered up to 100 ha, and served the Rhineland and Low countries much as modern Rotterdam does today.

Hamwic, Anglo-Saxon Southampton, was identified in much the same way when local people in Victorian times plundered the rubbish pits for their well-preserved bone and discovered silver coins at the same time. Major excavations since then have confirmed that in Wessex, as on the European continent, great trading centres were a feature of this age while old Roman towns lay in ruins. Hamwic, it now appears, covered an area of about 45 ha of low-lying ground beside the river Itchen. On the site of Ipswich, in the kingdom of the East Angles, there was another port of similar dimensions. Although unrecorded by the chroniclers of the age, Ipswich was undeniably a major centre, as excavations² over the past 10 years have shown.

Written sources describe other likely emporia, but until recently these have eluded archaeologists. Lundenwic (Anglo-Saxon London), according to the Venerable Bede, was a trading town of some importance by the late seventh century. Yet numerous excavations within the old Roman city indicate that the place was largely overgrown and abandoned. Recently, however, Martin Biddle and Alan Vince independently suggested^{3,4} that the Middle Saxon port lay not on the site of the Roman city, but in the Aldwych area west of the ruins. In 1986, excavations at Covent Garden, situated in this area, confirmed their hypothesis. Evidence of early mediaeval occupation, including buildings, pits and a great range of rubbish, worthy of a cosmopolitan rather than a rural community, was brought to light.

At York, virtually the same pattern has been discovered. Many excavations within the Roman and later Viking centre revealed little or nothing of the emporium described in writing by Alcuin, Charlemagne's Northumbrian minister. But excavations⁵ at the Redfearn glass foundry site, outside the east walls of the Roman and mediaeval city, revealed the characteristically rich debris of the port that had for so long been elusive. The settlement appears to have been located at the confluence of the rivers Ouse and Foss, though it remains to be seen whether it matches the other English dark-age towns in size and complexity.



Trading areas of seventh-century Europe. The inset shows the location of Quentovic.

Hamwic, Ipswich, London and York each seems to have served as a monopolistic centre for the Anglo-Saxon kingdom in which it was situated. As well as being a focus for long-distance trade, each was a regional centre for the production and marketing of commodities such as pottery, glassware, bone items and clothing. With the discovery of each new place it becomes possible to reconstruct the economic patterns of the period — matters which were of no interest to the contemporary monkish writers who passed through these places on pilgrimages and missions. Ceramic and glass debris, for example, can be related to the different groups of Frankish merchants operating in the North Sea area. Each new discovery makes it easier to reconstruct a phase in the development of towns in the time between the total breakdown of the urbanized economy of the Roman world and the rise of the well-documented markets of the tenth and eleventh centuries.

After Dorestad, Quentovic in northern France was perhaps the greatest centre north of the Alps in the seventh century. Both places were a springboard for Frankish trade with the English and Danes: Dorestad was the centre in which traders operating in the Rhineland, the

Low Countries and as far north as Denmark were based, whereas Quentovic seems to have been the base for traders covering northern and central France as well as southern England. Both ports were lynchpins in Charlemagne's ambitious economic strategy, becoming the focus of great wealth. Both inevitably attracted the attention of the Vikings, who repeatedly pillaged them⁶. Like Dorestad, Quentovic was deserted in the later ninth century, remaining a monument to a brief but influential phase in European urban history.

In 1984, a team from Manchester and Sheffield universities led by Dr David Hill began to search for this lost city in the Canche valley, east of Etaples. The Frankish sources offered some clues as to its whereabouts, but the time-consuming method of digging test pits turned out to be the only way to pinpoint its position. In 1985, the team identified it on low-lying ground in the commune of Vismarest at

the head of the valley of the river Canche, immediately below mediaeval and modern Montreuil and 15 km from the English Channel. Debris from the town is packed with pottery, bone, glass and industrial debris, and the test pits indicate that the town covered about 70 ha.

The first publication⁵ of some of these finds leaves no doubt that it is Quentovic that has been discovered. The pottery found confirms that the town had closer trading ties with the English than did Dorestad, for instance. It also shows that the port drew on a hinterland every bit as large as Dorestad, reaching southwards to the Paris basin and north as far as the Ardennes. The small investigations at Quentovic, like those in London and York, have to be treated with caution. Nevertheless, at Quentovic the occupation debris, left much as it was following Viking attacks in the 840s, remains tantalizingly undisturbed beneath pastureland. Peeling back the turf to reveal these deposits will be no less as important for historians of the dark ages than the discovery of Tutankhamun was for Egyptologists. At Quentovic, and in contrast to the English towns, there has been little robbing and no subsequent occupation. Indeed, it seems to have eluded discovery until now because its rich haul of animal bones, buried in the damp layers of the Canche valley, never held any attraction for the peasants of the Pas de Calais. □

1. Coutts, C. & Worthington, M. *Medieval Ceramics* 10, 23–28 (1986).

2. Hodges, R. *Dark Age Economics* (Duckworth, London, 1982).

3. Biddle, M. *Archaeology* 6, 23–27 (1984).

4. Vince, A. G. *Current Archaeol.* 310–312 (1984).

5. Kemp, R. *Interim Bull. York Archaeol. Trust*, 11, 8–16 (1986).

6. Lebecqz, S. *Ann. ESC* 2, 361–77 (1986).

Richard Hodges is in the Department of Archaeology and Prehistory, University of Sheffield, Sheffield S10 2TN, UK.

Lunar science

Impacts that magnetize

Mike Fuller

LUNAR magnetism has remained a fascinating enigma since the early Apollo missions recovered rock samples that carried strong natural remanent magnetism. The samples must have been magnetized in fields far stronger than present-day lunar fields, indeed at levels comparable to the Earth's magnetic field. Large-scale magnetization of the lunar crust was later detected with surface magnetometers and from orbit. Such magnetism was unexpected. Did the Moon once have a molten core that acted like a dynamo, or could transient fields at the surface of the Moon account for the magnetization? Clearly, the answer, if demonstrated unambiguously, could be important for both lunar and planetary science in general. L.L. Hood now suggests¹ an ingenious variant of the transient-field model to account for the location of crustal magnetization antipodal to young impact basins.

The main concentrations of crustal magnetic anomalies are near to the antipodes of certain young impact basins² and are in regions characterized by unusual pitted and grooved terrain, thought to be the result of convergent seismic energy generated by the impact. Hood suggests that impacts induced lunar plasmas that generated the necessary strong fields, and also created the conditions for magnetization to occur.

In arguing that a giant impact creates an expanding cloud of vapour and plasma, Hood follows other workers³, but makes the discussion more quantitative by estimating the mass of rock vaporized in an impact of the kind that made the Mare Imbrium. He also models the evolution of the cloud to estimate the sequence of changes in the surface field of the Moon.

A central aspect of the model is that as the cloud begins to expand around the Moon, field lines do not have time to diffuse out of the conducting mantle and will be compressed towards the antipode (see figure). Estimates of the conductivity of the lunar mantle, from electromagnetic-induction observations carried out during the Apollo programme, justify this assumption. An external field may be amplified by as much as 300 at the antipode. An initial field of a few nanotesla, typical of the present solar wind, should give a field at the antipode within an order of magnitude of the geomagnetic field.

Hood also notes that were there an ambient field from a lunar dynamo, the compression would give significantly larger surface fields. Although there are orders of magnitude of uncertainty in the

initial field assumption, such a mechanism could give sufficient field amplification near the antipode for even the weakest initial fields to generate significant fields. A difficulty, however, is that field line reconnection at the antipode may critically limit the amplification process (C. Russell, personal communication). Nevertheless, some amplification will take place.

For the material at the antipode to become magnetized, two key factors must be satisfied. The timescale of the magnetization process must be compatible with the decay timescale of the field; and the phenomenon that generates the magnetization must coincide with the presence of the field. According to the model, the field expands to the antipode in about 5 min. Hood estimates that the energy density in the plasma is sufficient to confine a compressed field for 10^4 – 10^5 s (for a few hours to a few days). The field will arrive before the seismic waves which take about 8 min (body waves) or 80 min (surface waves) to traverse the Moon. Ejecta on ballistic trajectories should take about 20–50 min. The critical question therefore remains: will the field decay too fast for relevant mechanisms of magnetization to function?

Thermoremanence, the acquisition of magnetization by hot, cooling rocks seems unlikely. The typical cooling time of large masses of buried rock does not match either limit of the estimated field-decay time, although even limited cooling can result in some magnetization. If the field lasts for a few days, however, substantial puddles of melt created by secondary impact of crustal fragments could be magnetized, as could smaller splashes.

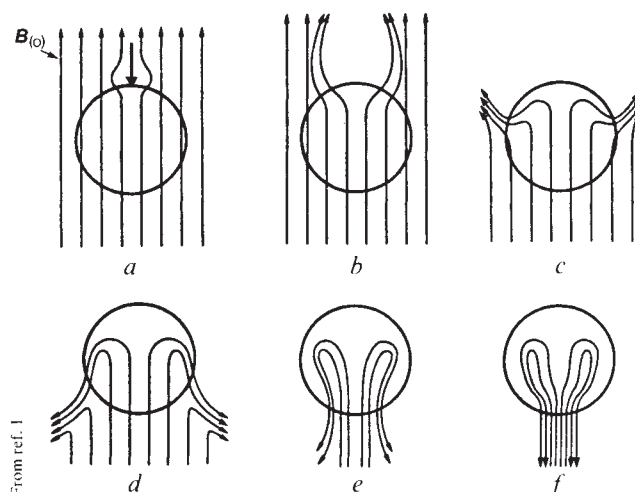
Shock-remanent magnetization is a much more plausible mechanism. There

are two possibilities to consider: first, magnetization caused by the passage of the seismic waves, associated with pressures estimated to be several hundred bars; second, shock magnetization associated with the secondary impacts of ejecta. Shock magnetization is not of much interest to terrestrial palaeomagnetists, but work on lunar magnetism⁶ suggests that given the proposed events at the antipode there would be substantial impact-related shock magnetization. Also, soil containing superparamagnetic and near superparamagnetic iron should be buried rapidly and become magnetized.

Any successful model, or combination of models, of lunar magnetism must explain the manifestations of lunar magnetism — the remanent magnetism of the samples; the magnetic fields observed on the surface; and those observed from orbit. It remains to be seen how Hood's model fits into such an overall scheme. Even though it was designed to explain the concentrations of crustal anomalies, it would be a more powerful idea if it also explained the other manifestations of lunar magnetism.

Unfortunately, the documentation of the magnetization of the samples is unsatisfactory. Most of the measurements were made before recent developments in palaeomagnetic techniques became established. The presence of native iron and iron-nickel as the principal magnetic carrier makes the recovery of ancient intensities more difficult than in the terrestrial case. Little consideration was given to the avoidance of magnetic contamination during collection or processing. Finally, the orientation of the samples at the time of magnetization is not known, so that tests of internal consistency cannot be applied. The problem is somewhat analogous to that of trying to elucidate the history of the Earth's magnetic field with evidence from about a hundred scree-slope samples.

Nevertheless, we may be reasonably confident of the following results. Samples that have natural remanent



From ref. 1

Generation of transient fields by a basin-forming impact (bold arrow). *a*, the Moon is initially in an ambient field $B(0)$; *b*, the impact creates a partially ionized silicate plasma which expands (*c*–*e*) around the Moon. The electrical conductivity of the Moon's interior prevents magnetic flux from diffusing out. The interaction with the ambient field compresses flux ahead of the cloud; *f*, field amplification, reconnection and diffusion provide transient enhancement of the field at the antipode.

magnetization indicative of magnetization in strong fields are about 3.9–3.6 gigayears (Gyr) old. The fields are smaller than, but of the same order of magnitude as, the geomagnetic field. With one possible exception, older and younger samples were probably magnetized in fields at least an order of magnitude smaller. Samples magnetized in strong fields include mare basalts probably carrying a primary thermoremanent magnetization acquired during the initial cooling of large masses of rock, and impact melt rocks from the highlands. The most strongly magnetic rocks are the breccias (shock-welded lunar soil and rock) that contain far more iron than the crystalline rocks. This is consistent with Hood's model. But the samples also suggest that Hood's model is not complete, because mare basalts that are not antipodal to young basins are also magnetized.

The surface magnetometer data⁵ and the orbital data from the fluxgate^{6,7} and back-scattered electron experiments⁸ reveal that the principal magnetic anomalies are in the highlands. In contrast, the lunar mares are generally magnetically subdued, with younger ones having weaker anomalies than the older ones.

Craters do not have particular magnetic expression. The orbital data also provide the key observation that the main crustal anomalies are roughly antipodal to four relatively young, large impact basins: Imbrium, Orientale, Serenitatis and Crisium. It is interesting that the ages of these basins (3.9–3.6 Gyr) are the same as those of the various rock samples that were probably magnetized in strong fields. It would have been a nice economy of theory had Hood's model explained all aspects of lunar magnetism, but it leaves some of them unresolved, such as the magnetization of the mare basins. □

1. Hood, L.L. *Geophys. Res. Lett.* **14**, 844–847 (1987).
2. Lin, R.P., El-Baz, F., Hood, L.L., Runcorn, S.K. & Schultz, P. *Lunar planet. Sci.* **XI**, 626–627 (1980).
3. Srnka, L.J. *Proc. 8th Lunar Sci. Conf.* 782–795 (1977).
4. Cisowski, S.M., Dunn, J.R., Fuller, M., Rose, M.F. & Wasilewski, P. *Proc. 5th Lunar Sci. Conf.* 2841–2858 (1974).
5. Dyal, P., Parkin, C.W. & Sonett, C.P. *Apollo 12, Preliminary Sci. Rep. NASA Spec. Publ.* **235**, 55–73 (1970).
6. Coleman, P.J., Lichtenstein, B.R., Russell, C.T., Sharp, L.R. & Schubert, G. *Proc. 3rd Lunar Sci. Conf.* 2271–2286 (1972).
7. Russell, P.J., Coleman, P.J., Soderblom, L.A., Stuart-Alexander, D.E. & Wilhelms, D.E. *Proc. 8th Lunar Sci. Conf.* 1171–1185 (1977).
8. Anderson, K.A., Lin, R.P., McGuire, R.E. & McCoy, J.E. *Space Sci. Instr.* **1**, 439–444 (1974).

Mike Fuller is at the Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA.

Evolutionary biology

Winning against history?

Paul H. Harvey and H.C.J. Godfray

In the opening address at a recent meeting*, A.J. Cain (Liverpool University) declared that "nature never evolved to be researched on, while humans never evolved to do research". If he is right, the ensuing talks demonstrated that evolutionary history has had little influence on current progress. In particular, we are no longer short of ingenious but testable explanations for the short-term evolutionary advantages of sex and genetic recombination.

Why is genetic recombination so widespread in nature? The 'red queen' explanation depends on continuously coevolving biotic interactions between competitors, pathogens and hosts or predators and prey. In support of this theory, G. Bell (McGill University) has demonstrated for the first time that cross-species comparisons within various taxa, including *Carex* sedges, seed plants, Orthopteran insects and mammals, reveal a positive relationship between one measure of chiasma frequency and the richness and intensity of negative biotic interactions such as parasitism, herbivory and predation. Such comparative tests do not exclude the possibility that other con-

foundings factors produce the relationship, but it is more difficult to conceive of these factors when the same relationships are demonstrated in diverse taxa.

Apparently allied to recombination, sex itself has always been a problem for evolutionary biologists. Two novel treatments of the topic were presented at the meeting. B. Levin (Amherst University) discussed the occurrence of sex and recombination in bacteria. Sex in this group is not coupled with reproduction and, with this point in mind, Levin asked how transformation, plasmid-mediated recombination and phage-mediated gene transfer evolved. He concluded that group-selection explanations are particularly appealing for the evolution of transformation in colonial bacteria. After excluding explanations involving parasitic DNA and individual selection, he argued that genetic recombination could have evolved coincidentally during phage- and plasmid-mediated gene transfer, possibly as a by-product of the need for DNA repair. In contrast to the situation in bacteria, sex in eukaryotes is coupled with reproduction, causing Levin to be sceptical about generalizing his arguments. This may be one of the few cases where what is true for *Escherichia coli* is not true for elephants. □

Where Levin's treatment divorces sex from reproduction, a new theory for the evolution of sex in eukaryotes divorces sex from recombination. A. Grafen (Oxford University) suggested that, if centrosomes replicate with error, an offspring might test whether the maternal or paternal centrosome is more efficient at microtubule initiation. The best centrosome could be chosen, resulting in non-mendelian inheritance. The ensuing best-of-two advantage could almost overcome the famous twofold cost of sex. Grafen gave clear criteria for testing his theory; the relevant information should not be long in coming.

When sexual reproduction has evolved, sex must be determined. Haplo-diploid sex determination is of particular interest because it is found in the hymenoptera, and has been implicated in the evolution of eusociality in this insect group. Under haplo-diploidy, unfertilized eggs develop into haploid males. An apparently functional equivalent occurs when males lose the paternal contribution to their genome during development. It has been thought that haplo-diploidy has the advantage over paternal genome loss because mothers can determine the sex of their offspring by deciding whether or not to fertilize their eggs (Bull, J.J. *Evolution of Sex-Determining Mechanisms* Cummings, Menlo Park, 1983). M. Sabelis (Leiden University) demonstrated that this theory cannot hold for phytoseiid mites, which have paternal genome loss, because he has shown experimentally that mothers can influence the sex ratio of their offspring.

From these considerations, two further questions arise. First, what is passed on to the egg which results in the eventual loss of the paternal genome? Sabelis and others suggest maternally transmitted enzymes or parasites. Second, why transfer the paternal genome at all? Sabelis speculates that the spare DNA could be used for repairing double-strand damage.

Despite the interest in sex and recombination, R.C. Lewontin (Harvard University) heralded the identification of DNA sequences as evolutionary biology's New Jerusalem. It is here, he argued, that the age-old controversy about the relative importance of identity by descent (evolutionary history) and identity by kind (selection) will be settled. But it was also with DNA sequences that Gabor Vida (Budapest University) warned British participants that their days on the international stage seem to be numbered. UGC, he pointed out, is just one mutational step from the stop codon. □

Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK, and H.C.J. Godfray is in the Department of Pure and Applied Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK.

*Founding Congress of the European Society for Evolutionary Biology organized by S.C. Stearns and A.J. van Noordwijk. Basel, Switzerland, 26–30 August 1987.

Quantum physics

Quantum phase and magnetic flux

Ulrich Eckern

THE Aharonov–Bohm effect, first described in 1959, is one example of the mysterious properties of quantum mechanics. Charged particles, propagating around a magnetic field in field-free vacuum, acquire a phase shift of their wavefunctions that has observable consequences. This has been confirmed — again, but now incontrovertibly — in a recent experiment¹. Even more surprising, however, is the appearance of Aharonov–Bohm oscillations in some non-superconducting solid-state devices^{2–5} which, as

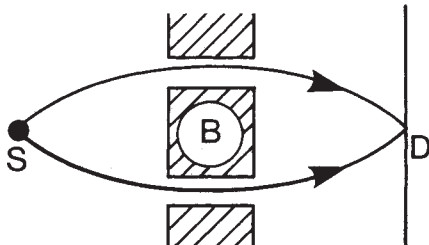


Fig. 1 The idealized two-slit geometry for the Aharonov–Bohm effect. An electron beam from source S passes through two slits to arrive at the detector D. The magnetic field, which points into the page, is confined to the shaded regions. Despite this, it still affects the quantum-mechanical phase of the electrons.

well as indicating possible technical applications⁴, contributes significantly to our understanding of quantum interference on a macroscopic level. The most recent example, from G. Timp *et al.*⁵, demonstrates that the effect occurs in high-mobility semiconductor samples.

In the typical (idealized) geometry (Fig. 1), an electron beam travels from a source to a detector by two different paths; the magnetic field (**B**) is confined to a long solenoid threading a region, inaccessible to the electrons, between the paths. The interference pattern thus detected is a periodic function of the magnetic flux, Φ , with the period given by $\Phi_0 = hc/e = 4.135 \times 10^{-7}$ gauss cm², although the electrons never enter into a region of non-zero field. Alternatively, one can imagine an electron confined to a box that is transported very slowly (adiabatically) on a closed loop around the flux⁶. According to the laws of quantum mechanics, the phase, φ , of the electron wavefunction changes by $2\pi/\Phi_0$ times the line-integral of the vector-potential, **A**, along the loop. (A is to magnetic field roughly as electrostatic potential is to electrostatic field.) Thus $\Delta\varphi = 2\pi\Phi/\Phi_0$. These and similar considerations point to the fundamental role that electromagnetic potentials, rather than the fields, play in quantum physics of charged particles.

In one recent experiment¹ that was

designed as a critical test of the effect, a tiny toroidal magnet was covered entirely by a superconductor (niobium), to prevent any leakage of the magnetic field into the electron path. The electron beam passing through the ring and the reference beam were brought together by an electron biprism to form an interference pattern, which was enlarged 1,000 times by electron lenses and recorded on film as a hologram. Because the magnetic flux enclosed by a superconductor is quantized in units of $\Phi_0/2$ (see the recent News and Views article by David Caplin⁷), the phase difference between the two beams is restricted to $\Delta\varphi = 0$ or π (modulo 2π). A typical result, with an odd number of flux quanta, is shown in Fig. 2.

In the past few years, the Aharonov–Bohm effect has also been detected in various non-superconducting solid-state devices, including metal cylinders² and rings³, and in semiconductor microstructures^{4,5}, provided that the structures are small ('mesoscopic', scale $\sim 1 \mu\text{m}$) and the temperature is low ($< 1 \text{ K}$). In a typical experiment, small oscillations of the resistance are recorded as an applied perpendicular field is varied, with a dominant period corresponding to Φ_0 in thin rings, and $\Phi_0/2$ in the cylinder geometry.

Most results can be explained using Feynman's 'path summation' picture. Applied, for example, to a metal ring (see Fig. 3) with a diameter much larger than

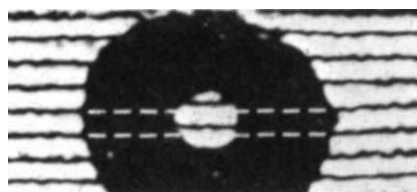


Fig. 2 Interference micrograph of a toroidal magnet¹. The interference fringes show that electrons passing through the middle of the magnet have had their phase shifted by π relative to the electrons outside the rings — otherwise the fringes would be aligned.

the metal width (so that the enclosed flux is well defined), this leads to the conclusion that the resistance of the device varies according to whether the electrons passing through the two semicircles recombine in phase, in antiphase or in between. Thus the resistance has a component that is periodic in Φ , with period Φ_0 . Because of imperfections and impurities, one might expect the effect to be unobservable. But elastic scattering does not destroy quantum coherence, unlike inelastic processes such as scattering by phonons; the latter, fortunately, are negligible for mesoscopic

samples at very low temperatures. On the other hand, because of the random impurity distribution, it is important that the phase difference to start with (at zero field) is also a random quantity.

Thus, the Aharonov–Bohm oscillations, as well as the slow random variations onto which they are superimposed⁸, have to be considered as fingerprints of an individual sample. An average over many samples, as done effectively in the cylinder geometry², completely washes out the Φ_0 oscillations, leaving only the $\Phi_0/2$ oscillations. The latter, in fact, arise from a

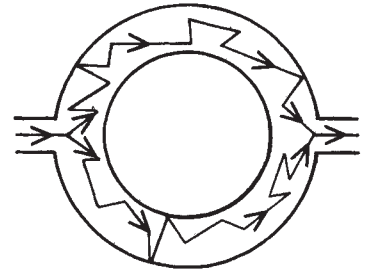


Fig. 3 Motion of electrons through a metal ring. Elastic collisions off impurities do not destroy quantum coherence. Electrons take a more direct, or ballistic, path through high-purity, high-mobility semiconductors.

higher-order interference of paths that encircle the magnetic flux, with their time-reversed counterparts⁹.

In their recent article⁵, Timp *et al.* report the first unambiguous observation of the Aharonov–Bohm effect in semiconducting rings of interleaved GaAs and AlGaAs ('heterostructures'). In contrast to an earlier experiment⁴ on similar device that saw only about three periods of the oscillation, the new experiment shows pronounced oscillations that persist up to $\Phi \sim 300\Phi_0$, with a period that can only be associated with Φ_0 .

That the oscillations are so strong, accounting for 10 per cent of the resistance is perhaps not so surprising, as the samples have very high charge mobility; the distance travelled between scattering events is comparable to the size of the device. Under favourable conditions, the variations are expected to be even stronger. Thus the experiment also indicates the possibility of potential applications in rapid-switching devices⁴ that modulate resistance with a transverse electric field using the electrical Aharonov–Bohm effect. This effect, however, remains to be demonstrated experimentally. □

1. Tonomura, A. *et al.* *Phys. Rev. Lett.* **56**, 792–795 (1986).
2. Sharvin, Yu. V. *Physica* **126B**, 288–293 (1984).
3. Webb, R. A. *et al.* *Phys. Rev. Lett.* **54**, 2696–2699 (1985).
4. Datta, S. *et al.* *Phys. Rev. Lett.* **55**, 2344–2347 (1985).
5. Timp, G. *et al.* *Phys. Rev. Lett.* **58**, 2814–2817 (1987).
6. Berry, M. V. *Proc. R. Soc. A* **392**, 45–57 (1984).
7. Caplin, D. *Nature* **326**, 827–828 (1987).
8. Lee, P. A. *Nature* **324**, 613–615 (1986).
9. Chakravarty, S. & Schmid, A. *Phys. Rep.* **140**, 193–236 (1986).

Ulrich Eckern is in the Institut für Theorie der Kondensierten Materie, University of Karlsruhe, Postfach 6980, D-7500 Karlsruhe, FRG.

Earthquake damage in Mexico City

SIR—Flores *et al.* conclude from their study of resonance and earthquake damage in Mexico City that the surface distribution of damage within the soft-ground area can be explained by destructive and constructive interference of waves in the soft foundation material of the extinct lake¹. During an extended visit to the area I was first struck by 'pockets of damage' scattered over the soft-ground region, and entertained similar ideas about interference of waves.

A detailed inspection of all buildings of more than five storeys, however, produced a far simpler explanation of the phenomenon. Tall buildings are not uniformly distributed over the Lago de Texcoco area of Mexico City but concentrated in regions like Roma Norte, Tlatelolco and Reforma. Predominantly these buildings suffered because of what may be called resonance². Flores *et al.* also mention selective destruction of tall buildings. Buildings of few storeys generally suffered very little or no damage even if old and vulnerable. In the areas with low or no damage there were practically no tall buildings and therefore no structures likely to be damaged.

We do not deny the possibility of constructive or destructive interference. Earthquake damage is, however, controlled by many parameters³ and it is therefore very dangerous to rely on theoretical models only, especially if they are not supported by detailed observations. It must be remembered that earthquakes cause untold misery and inflict colossal financial losses. This is even more reason to compare carefully results produced by any model with the data derived from as large a sample as possible. This time-consuming and costly sampling is dictated by the many damage parameters which must be considered. Some examples may serve to illustrate the potential pitfalls and the risk from future earthquakes if such rules are not heeded:

- The boundaries of the soft-ground region are not well known in the eastern region of the extinct lake. Any theoretical calculation is therefore bound to incorporate considerable uncertainties.
- The density of the soft material and therefore the wave velocity increases with depth. As the composition of the soft material is not homogeneous a very complex model and more precise information than available today would be needed to calculate interference patterns reliably.
- The heterogeneous soft material is excited by a very complex pattern of waves arriving from the rupture plane. Also this causes interference of waves in the soft material, adding to the many uncertainties.

- The comparatively few accelerographs did not give a uniform predominant period of 2 s but values ranging between ~1.5 and 3 s. One may therefore wonder whether Flores *et al.* were right to select among the many solutions obtained only those which have the 'correct' frequency.

- Hydrologists or meteorologists know that adding rain gauge data changes observed rainfall patterns. How reliable is the distribution of ground shaking derived from only a handful of instruments?

- Damage to buildings depends predominantly on: their structural strength; the effect of stiffening walls; architectural layout such as degree of irregularity of floor plan and of elevations (L-shape, U-shape, H-shape, T-shape); uniformity of distribution of masses, stiffness and damping; compatibility of building materials; orientational sensitivity; site effects (resonance). The importance of these parameters has been seen after many earthquakes⁴⁻⁸. Flores *et al.* have not considered these parameters in their conclusions.

Interesting as interference patterns may be, it appears to be dangerous to explain the damage pattern observed in Mexico City by the interference of waves in the soft ground.

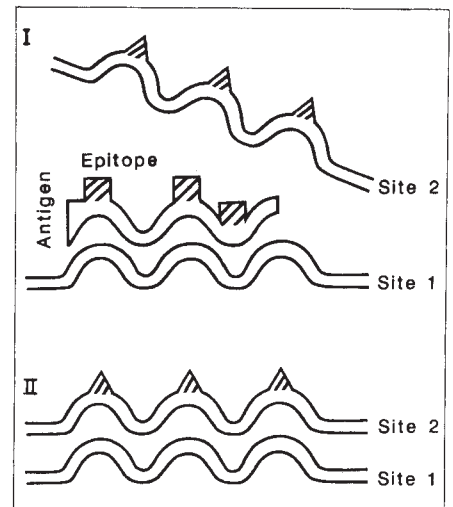
HERBERT TIEDEMANN

Swiss Reinsurance Co.,
CH-8022, Zurich,
Switzerland

1. Flores, J., Novaro, O. & Seliman, T.H. *Nature* **326**, 783-785 (1987).
2. Tiedemann, H. *Eighth Symp. Earthquake Engng* **1**, 23-30 (University of Roorkee, 1986).
3. Tiedemann, H. *Engineering Geology* **20**, 169-179 (1984).
4. Inai, Ts. & Okubo, T. *Seventh World Conf. Earthquake Engineering* Vol. 3, 33-40 (1980).
5. Marsal, R.J. *Symp. Org. Nac. de Est. Ing. Ingenieria* **28**(1), 10-23 (1957).
6. Rosenblueth, E. *Ingenieria XXXI*(1), 1-31 (1961).
7. Shibuya, J., Kimura, H. & Shiga, T. *Seventh World Conf. Earthquake Engng* Vol. 2, 199-206 (1980).
8. Tiedemann, H. *Seventh World Conf. Earthquake Engng* Vol. 2, 267-270 (1980); Vol. 6, 617-624 (1980); *Seventh Eur. Conf. Earthquake Engng* Vol. 6, 27-34 (1982); *Eighth Eur. Conf. Earthquake Engng* Vol. 2, 4.1.63-4.1.66 (1986).

Does the T-cell receptor bind to the MHC?

SIR—I would like to point out one very important consequence of the model proposed by Schwartz¹ and Guillet *et al.*² of the molecular interactions between major histocompatibility complex (MHC) molecules and antigen. The present model postulates that the T-cell receptor interacts with both antigen and MHC molecules. There is now strong experimental evidence for the interaction between antigen and MHC, but evidence for interaction between receptor and MHC is circumstantial. The idea stems from the classical concept of dual recognition, itself derived from the phenomenon of T-cell restriction. This phenomenon does not itself require that a direct interaction between receptor and (presumably)



The two configurations of the MHC molecule.

polymorphic elements of the MHC exists, as a particular MHC molecule may select a unique portion of antigen, imposing a unique configuration on it³. The form of the epitope presented to T cells will therefore be different for different MHC molecules and this will result in the T cells being restricted to one MHC molecule. But if we take this view of restriction, it is very difficult to see how the peripheral T-cell pool can develop a repertoire which preferentially recognizes the MHC, either 'self' or foreign, without directly interacting with it. The evidence for such a skewed repertoire is now strong, if not conclusive (see ref. 4 for review).

The model proposed by Schwartz and his colleagues offers an easy solution to this problem. As shown in the figure, the MHC molecule exists in two configurations: in one the a and b chains (of Class II molecules) are prised apart, and an antigen binds to site 1, where it is recognized by a T cell. In the other configuration, the antigen is replaced by a site 2 of the MHC molecule. The crucial feature is that antigen and site 2 show molecular similarities, perhaps reflecting the fact that both bind to site 1. T-cell receptor recognition of self-MHC plus antigen then results from recognition of those features of an antigen that mimic the self-MHC molecule. To explain why the development of the repertoire is dictated by the MHC of self, I suggest that only those T cells which recognize key polymorphic features on site 2 of the MHC (in closed position, shaded area in part II of the figure) are selected. To bind and trigger the T-cell receptor, an antigen bound to site 1 will have to share some structural features with site 2, with differences constituting the T-cell epitope which presumably increase the affinity of the antigen/receptor interaction enough to result in T cell activation. In this model, as opposed to that of Schwartz, the similarities between antigen and site 2 are dictated both by the

requirement that both bind to site 1, and by the T-cell repertoire itself. This would mean that the T-cell receptor does not bind to MHC simultaneously with antigen, but recognizes certain similarities between antigen and MHC as on the altered-self model^{5,6}. I suggest that there is no need to postulate direct interaction of receptor and MHC, with all the conceptual difficulties this has entailed. Indeed, this model does away with the whole conception of classical 'dual recognition'. The T-cell receptor, just like antibody, binds to antigen alone. But because the structure of antigen is both selected and determined by the MHC, in the antigen the T-cell receptor can be said to 'see' a ghostly reflection of the MHC.

I am grateful to Professor Mitchison and other colleagues for their advice and criticism which have helped to clarify the issues involved.

BENJAMIN M. CHAIN

Department of Zoology,
University College London,
Gower Street,
London WC1E 6BT, UK

1. Schwartz, R. *Nature* **326** 738–739 (1987).
2. Guillet, J.-G. *et al. Science* **235**, 865–870 (1987).
3. Weidelin, O. *Immun. Today*, **8**, 80–84 (1987).
4. Bevan, M.J. in *Lymphocytes Today* (ed. Inglis, J.R.) 38–41 (Elsevier, Amsterdam, 1983).
5. Matzinger, P. *Nature* **292**, 497–501 (1981).
6. Kourilsky, P., Chataat, G., Rabourdin-Combe, C. & Claviere, J.-M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3400–3404 (1987).

Growth regulating network?

SIR—A recent article by W.J. Rutter and co-workers¹ suggests that the receptor for insulin-like growth factor II (IGF-II) may be a carbohydrate-recognizing protein. This prediction comes from the finding that the complementary DNA sequence for the human IGF-II receptor is very similar to that of a bovine carbohydrate-binding protein² which recognizes 6-phosphorylated mannose and is believed to be involved in the routing of glycosylhydrolases to lysosomes.

This news comes at a time when an increasing body of immunochemical and other biochemical observations (reviewed in ref. 3) has implicated the carbohydrate structures of glycoproteins, glycosaminoglycans and glycolipids in cell growth and differentiation processes. One particular set of observations (reviewed in ref. 4), namely, the stimulation of autophosphorylation of the solubilized epidermal growth factor receptor by perturbing its carbohydrate moiety with antibody, has led us to propose⁴ the existence *in vivo* of growth-regulating (growth-tuning) networks based on protein-carbohydrate interactions and involving growth factors and their receptors.

Is the IGF-II receptor one component of such a network of growth regulators? More specifically, is the IGF-II receptor a

bi- or multifunctional protein with one combining site for IGF-II and another for carbohydrate structure(s) on other growth-regulating glycoconjugate(s) that form a network which involves the insulin and IGF-I receptors? Such a mechanism of intracellular communication could explain the IGF-II effects that seem to be mediated by the receptors for insulin and for IGF-I, and it would have an in-built fine tuning element dependent on glycosylation.

TEN FEIZI

ROBERT A. CHILDS

MRC Clinical Research Centre,
Harlow HA1 3UJ, UK

1. Morgan, D.O. *et al. Nature* **329**, 301–307 (1987).
2. Lobel, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 2233–2237 (1987).
3. Feizi, T. *Nature* **314**, 53–57 (1985).
4. Feizi, T. & Childs, R.A. *Biochem. J.* **245**, 1–11 (1987).

Strand segregation or recombination

SIR—A recent *Nature* article by A. J. S. Klar¹ concluded that the segregation of modified parental DNA strands in the yeast *Schizosaccharomyces pombe* determines which daughter cell undergoes mating-type switching. Klar proposes that a DNA modification at the *smt* locus is necessary for the double-stranded break which initiates switching. This modification is assumed to exist on only one DNA strand, thus it is inherited by only one daughter cell. The evidence for this hypothesis is the *in vivo* existence of an excised DNA fragment from a strain carrying a direct repeat of the *mat1* (including *smt*) locus, while a strain carrying an inverted repeat does not produce an excised fragment. Klar concludes that this results from the co-existence of two active cleavage sites only in cells with the direct repeat.

An alternative explanation is that homologous recombination between the direct repeats of *mat1* can result in the excision of this same region of DNA as a circular fragment. Cleavage of this circle at its single *smt* site would produce the same linear fragment observed by Klar. In addition, the band observed migrating more slowly than the linear fragment during gel electrophoresis of undigested DNA (which is not explained by Klar's model) could represent the relaxed circular excision product. Only the direct repeat of *mat1* will result in excision by homologous recombination. Recombination at the inverted repeats will merely invert the region between the repeats. The apparent lack of chromosomes with two *smt* cleavages in cells with the inverted repeat can be explained by assuming that only one cut can occur per chromosome (for example, if supercoiling is required).

Although excision by homologous recombination at direct repeats normally occurs at low frequencies², the formation of a double-stranded break at the region

of homology (which occurs *in vivo* at the *smt* locus) could greatly stimulate the process³. The analogous insertion of DNA by homologous recombination in *Saccharomyces cerevisiae* is stimulated 10–1,000-fold by double-stranded breaks⁴. Indeed, circular excision products formed by recombination between the direct repeats of *mat1* and the silent *mat* cassettes have been observed previously⁵, showing that the process occurs frequently enough to be detected.

As parental strand segregation could account for many aspects of development, it will be of interest to determine whether this, or homologous recombination, is responsible for Klar's results. The question seems experimentally approachable either by conclusively identifying the circular excision product or detecting the predicted recombination products of the inverted repeats through appropriate restriction digests and Southern blots.

DAVID W. RUSSELL

Laboratory of Genetics,
The Rockefeller University,
1230 York Avenue,
New York, New York 10021, USA

1. Klar, A.J.S. *Nature* **326**, 466–470 (1987).
2. Jackson, J.A. & Fink, G.R. *Nature* **292**, 306–311 (1981).
3. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. & Stahl, F.W. *Cell* **33**, 25–35 (1983).
4. Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6354–6358 (1981).
5. Egel, R., Beach, D.H. & Klar, A.J.S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3481–3485 (1984).

KLAR REPLIES — The generation of a 6.1-kilobase fragment in a strain with a direct *mat1* duplication as a consequence of intrachromosomal recombination between direct repeats of *mat1* is unlikely because the strains we used, as stated in our paper¹, are recombination deficient because of their *swi5*⁻ genotype. The *swi5*⁻ strains are defective in the utilization of the double-stranded break at *mat1* for recombination^{2,3}, and are known to be defective in meiotic general recombination at the *ade6* locus^{4,5}. Thus, I think that Russell's interpretation is incompatible with the presented results.

AMAR KLAR

Cold Spring Harbor Laboratory,
Cold Spring Harbor, New York,
New York 11724, USA

1. Klar, A.J.S. *Nature* **326**, 466–470 (1987).
2. Egel, R., Beach, D.H. & Klar, A.J.S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3481–3485 (1984).
3. Klar, A.J.S. & Miglio, L.M. *Cell* **46**, 725–731 (1986).
4. Gutz, H. & Schmidt, H. *Curr. Genet.* **9**, 325–331 (1985).
5. Schmidt, H., Kapitzka, P. & Gutz, H. *Curr. Genet.* **11**, 303–308 (1987).

Sickle cell detection: Erratum

IN the Scientific Correspondence "Detection of sickle cell anaemia and thalassaemias" (*Nature* **329**, 293; 1987), ref. 8 should have been cited at the end of the second paragraph. The paper, by S. Kogan, M. Doherty and J. Gitschier, has been published in the 15 October issue of *New England Journal of Medicine*. □

What's bred in the bone

Leonard Krishtalka

Laetoli: A Pliocene Site in Northern Tanzania. Edited by M.D. Leakey and J.M. Harris. Clarendon: 1987. Pp.591. £90, \$150.

LAETOLI, in northern Tanzania, is the Masai name for the local red lily. Mary and Louis Leakey were first drawn to the place of the red lilies in 1935, when a Masai visitor to their camp at Olduvai Gorge told them of an area to the south where there were more "bones like stone". Sensing their scepticism, the Masai, named Sanimu, left on foot for Laetoli, about 30 km away, and returned with a handful of fossil bones and teeth. The Leakeys quickly decamped to Laetoli and found more mammalian material, which eventually was given to the British Museum (Natural History). Unrecognized among the Laetoli fossils was a primitive hominid canine, which was catalogued as that of a cercopithecoid monkey. Also unrecognized was the antiquity of the fossiliferous beds, then thought to be equivalent to Olduvai Bed I and II, or 1–2 million years (Myr) old.

Such inauspicious beginnings belied what was to become one of Africa's most spectacular hominid localities. *Laetoli: A Pliocene Site in Northern Tanzania* is its chronicle: a magnificent monograph, comprehensive in scope, rich in detail and elegant in presentation.

After 1935, Laetoli lay fallow palaeontologically for 40 years, save for a foray in 1939 by L. Kohl-Larsen, a German explorer, who recovered some fossil remains, including some of hominids, and brief collecting trips by the Leakeys and co-workers in 1959 and 1969. Then, in 1974, just two years after Louis's death, another visitor to Olduvai rekindled palaeontological interest in Laetoli. George Dove, a friend and neighbour of the Leakeys, had found fossil antelope and horse teeth in a load of sand hauled from Laetoli for building material. Mary Leakey's ensuing expeditions there that year struck geological and anthropological paydirt: they found primitive hominid jaws and teeth in the Upper Laetoli Beds (the Anglicized and formal geological variant of Laetoli), which were dated as more than 2.4 Myr old, the age of an overlying lava flow. The Laetoli hominids predated those from Bed I, the oldest horizon at Olduvai.

Indeed, two years later, radiometric dating of the volcanic Laetoli Beds showed these wind-deposited and 'airfall' ash layers to represent about 300,000 years of Pliocene time between c. 3.5 and 3.8 Myr ago. One of these ash falls records only a few hours of Pliocene time during which

three hominids, perhaps a male, female and child, strolled across the place of red lilies, leaving two parallel trails of their footprints in the soft, damp ash. Other animals, from insects to gazelle to elephants, also left their prints in the ash layers of the Laetoli Footprint Tuff, almost 10,000 prints at nine sites. Successive eruptions of Sadiman, a volcano



Steps in time — the hominid footprints at Laetoli.

35 km east of the Laetoli area, buried and preserved the Tuff, its footprints and some of the bones of the animals who made them.

The hominid trails, discovered in 1978, were dramatic evidence that the small-brained hominids who meandered across the Laetoli savannah 3.5 Myr ago did so with a modern human gait. Clearly, acquisition of an upright, fully bipedal posture had long preceded cranial enlargement in hominid evolution, a conclusion that had already been drawn from 'Lucy', the remarkable hominid skeleton discovered in 1972 by Donald Johanson at Hadar, Ethiopia.

Laetoli the book records the eight years (1974–1981) of multidisciplinary field work at Laetoli carried out by Mary Leakey and her colleagues. Their research encompasses the geology, geochronology, palaeontology and palaeobotany of the entire Laetoli succession (Laetoli Beds, Ndolyana Beds, Ogoi Lavas, Olpiro Beds, Ngalopa Beds, from oldest to youngest). In 14 chapters and four appendices, 33

contributors describe in exhaustive detail the fossil localities, the history and logistics of the field work, the geology and dating of the sediments, the fossil and Recent plants, and the fossil reptiles, birds, insectivores, primates, rodents, lagomorphs, carnivores, aardvarks, proboscideans, perissodactyls, artiodactyls, termitaries, bees, molluscs, and animal and hominid footprints. The editors themselves contribute eight chapters or parts thereof, four by John Harris (deinotheres, suids, giraffids and camelids, summary) and four by Mary Leakey (Introduction, summary of hominid remains, animal prints, hominid footprints).

Laetoli is one of those rare edited works in which the contributions are consistently excellent. Each of the faunal and floral treatments includes the systematics, description, meristics, illustrations and affinities of the material, and, where possible, the implications of the taxa for African Pliocene biostratigraphy and palaeoenvironmental reconstructions. The sections on the history of field work and on the localities, geology, geochronology, animal and hominid footprints, and overall summary are equally thorough.

What emerges from these chapters is a clear view of the palaeoenvironment of the Laetoli hominids, a finer biostratigraphical and chronostratigraphical resolution of the East African Pliocene, taxonomic and phylogenetic revisions (especially among the reptiles, rodents, monkeys, elephants and artiodactyls) and excellent grist for milling the evolutionary patterns through the Laetoli succession and among East African Pliocene assemblages. The Lower Laetoli Beds (c. 4.3–3.8 Myr old) preserve over 21 species of vertebrate, including rare aquatic forms (fish, crocodile) but no hominids. Otherwise, the fauna is comparable to, though less diverse than, that from the Upper Laetoli Beds (c. 3.5–3.8 Myr old), which includes over 97 species, 15 of them new. The upper unit of the Ndolyana Beds (3.5–2.4 Myr old) yielded 44 species of invertebrate and vertebrate, one of them new. None of the horizons preserve cultural artifacts. Unlike other Pliocene sites in the Eastern Rift Valley, the Laetoli area was a dry savannah much of the time (from 3.8 Myr ago), with no large lakes or rivers.

Ironically, missing from *Laetoli* is a description of the Laetoli hominids. Leakey explains in Chapter 5, "The Laetoli Hominid Remains", that T.D. White granted and then withdrew his permission for the editors to reprint his earlier descriptions of the hominids published in 1977 and 1980 in the *American Journal of Physical Anthropology*. As a result, Chapter 5 is succinct: a fine, five-page photographic record of the hominid material, and a two-page summary and

John Reader/Mary Leakey

table of their morphology and provenience.

The absence of White's articles is more a matter of convenience than substance. Readers intent on marrying the photographs of the hominids to the descriptions can obtain the relevant issues of the journal. Those intent on a different phylogenetic analysis of the hominids will have to wait. In her Introduction, Leakey strongly questions the systematic philosophy and taxonomic practice behind *Australopithecus afarensis*, the new species erected in 1978 by Johanson. White and Yves Coppens for all of the Laetoli and Hadar hominid material. But her treatment of the hominid issues, she hints, will come later.

If the hominid bones are missing, their footprints are not. Chapter 13 describes the excavation of the three hominid trails (G-1, G-2, G-3) and the 69 individual prints (M.D. Leakey); two independent studies (L. Robbins, R. Tuttle) reconstruct the foot morphology, stature, stride and gait of the three hominids. Trails G-2 and G-3 form a single overprinted trail, the result of the G-3 hominid having followed closely and deliberately in the footsteps of the G-2 trailblazer. According to Robbins, the prints are of human form; their size implies that G-1 was about 117 cm tall, G-2 176 cm and G-3 147 cm, or within the range of modern American children, adult males and adult females, respectively. The temptation to speculate that the trails were left by a strolling family is acknowledged and resisted.

Tuttle is more decisive. His analysis ranges from the Laetoli trails to the footprints of Solomon Islanders and Californian students. He concludes that the Laetoli print-makers would pass for *Homo*, even *H. sapiens*, were *Homo* defined on footprints alone — footprints 'Lucy' and her kin from Hadar could not have made, given the anatomy of their foot bones. As such, Tuttle buttresses Leakey's obvious scientific disdain for *A. afarensis* as well as her view that the Hadar and Laetoli hominid material should comprise different and diverse species.

The only serious omission from the book is a palaeoecological synthesis, dealing with the trophic structure and guilds of the Laetoli fossil communities. However, the chapters brim with data for such an analysis, which, one hopes, will be forthcoming.

Unlike the typical symposium smorgasbord, *Laetoli* is a superbly edited opus, cohesive in subject and scholarship. Its editors have negotiated this terrain before in monographs on Koobi Fora and Olduvai Gorge, and here continue their extraordinary contributions to the Plio-Pleistocene palaeobiology of Africa. □

Leonard Krishtalka is Associate Curator in the Section of Vertebrate Fossils, Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA.

Academia meeting industry

James A. Barnett

The Yeasts, 2nd Edn, Vols 1 and 2. Edited by A.H. Rose and J.S. Harrison. Academic: 1987. Vol. 1 pp.423, \$69.50, £38.50. Vol. 2 pp.309, \$59.55, £33.

WITH few of the original authors remaining, and so only nominally a second edition, this is a completely new survey of work on yeasts, currently classified into about 60 genera and 500 species. Yeasts are unicellular fungi which reproduce vegetatively by budding or fission, but some are stages of hyphal fungi. They are highly versatile organisms and, although only very few species have been exploited so far, many are particularly suitable for industrial purposes because they are often easier to handle than bacteria or moulds.

Volume 1 of *The Yeasts* reviews taxonomy, ecology, human pathogens and also the cell cycle (of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*). Volume 2 considers nutrition and inhibitory compounds, temperature relations, effects of radiation, yeast cultivation, killer yeasts, flocculation, sexual agglutination and adhesion to (mainly) biological surfaces. Organelles and physiology will be dealt with in Vols 3 and 4, yet to be published; two further volumes are planned, presumably covering genetics and industrial applications.

Although the subjects are something of a hotchpotch, they all relate to recent research. Indeed, even in the chapter on yeast classification, more than 60 per cent of the references were published after 1970. So, despite some cross-referencing, the chapters are in some ways more akin to the contents of a review journal than to a text book. Accordingly, most chapters will interest those already at least partly familiar with the subject matter but would make rather difficult reading for many students of general microbiology.

Three chapters concern current problems in yeast taxonomy. One such problem stems from the tacit assumption that classification must show ancestral relationships; since there is no useful fossil record, these relationships can be assessed only in terms of the amount of genetic similarity and difference, not genealogy. Overall genetic differences are probably best measured by DNA hybridization. However, some such measurements have caused consternation, because ability to cross is a well-established criterion for belonging to the same species and some sexual yeasts with low DNA similarity have crossed successfully and, conversely, some with high similarity have failed to

cross. Clearly, a small number of mutational changes might include one affecting sexual reproduction. On the other hand, genes affecting sex may remain unaffected, whilst many others have mutated. Therefore, these modern techniques, such as comparisons of nucleic acids, do not enable taxonomists to provide a more precise concept of species than that of Darwin in the first edition of *The Origin of Species*. Yet taxonomic precision and ease of identification are important, as is obvious in the context of ecological surveys and studies of the pathogenicity of yeasts.

The practical implications of microbiology are apparent, even in this primarily academic work on yeasts. Many ecological studies have been supported because of the role of yeasts in food spoilage and in human and animal diseases. Yeasts are ubiquitous and have a wide range of activities. For example, some yeasts can ferment and spoil honey, a remarkable achievement physiologically, as honey is composed of 70–80 per cent of hexoses. Yeasts pathogenic to man may occur in skin, mouth, vagina, blood or brain. Work on the yeast cell cycle concerns the mechanism of coordination between cellular growth and cell division, so that division occurs when the cell reaches a particular size. Such studies on simple eukaryotes should, in the long run, be important for understanding human growths.

The reviews in this work contain a good deal of practical, methodological detail. Overall they are authoritative and provide essential material for those teaching microbiology and researching on yeasts or developing them for industrial use. □

James A. Barnett is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7UJ, UK.

Erratum

In the review of *Yeast* (new journals' review issue, *Nature* 329, 373; 1987), the subscription price for countries other than Britain should have read \$110, not £110.

New in paperback

- *Genetic Takeover and the Mineral Origins of Life* by A.G. Cairns-Smith. Publisher is Cambridge University Press, price is £15, \$24.95. For review see *Nature* 300, 127 (1982).
- *Seeds of Change: Five Plants That Transformed Mankind* by Henry Hobhouse. Publisher is Harper & Row (under the Perennial Library imprint), price is \$8.95. For review see *Nature* 319, 21 (1986).
- *Ada: A Life and a Legacy* by Dorothy Stein (a biography of Augusta Ada Byron, later Countess of Lovelace and helpmeet of Charles Babbage). Publisher is MIT Press, price is \$9.95, £8.95. For review see *Nature* 320, 224 (1986).
- *The Nazi Doctors: A Study in the Psychology of Evil* by Robert Jay Lifton. Publisher is Macmillan, London (under the Papermac imprint), price is £9.95. For review see *Nature* 323, 677 (1986).

Childish concerns

P. E. Bryant

A Good Enough Parent: The Guide To Bringing Up Your Child. By Bruno Bettelheim. *Thames & Hudson/Knopf*: 1987. Pp.377. £12.95, \$18.95.

THE relationship between science and practical necessity is often an uneasy one. People had to get on with the job of building houses and bridges, treating the sick, managing the economy and bringing up children long before there was a scientific rationale for what they were doing. In some areas our scientific knowledge has supplanted the practical expertise which preceded it by so many centuries; we now know, for example, what makes bridges stand up and why some medicines work.

Other subjects have not reached this fortunate and exact state. It is not at all clear yet what makes an economy prosper, nor why some children have a calm and happy childhood and others do not. Yet the practical problems still have to be tackled. The result is that in many subjects, of which of course child psychology is one, there are at least two levels of discourse about the same questions. On the one hand we have the wise men and women, often with many years of experience in clinics or classrooms, who have to make immediate decisions about how to solve a wide range of children's problems. On the other there are the scientists struggling, often to no avail, with the extremely complicated matter of cause and effect in children's emotional and intellectual growth.

What should we be hearing from these two sets of people? At first sight the answer seems quite straightforward. The practitioners should be telling us: "Look, it seems to me that when I do A with young children the result is usually B; so if you also want to achieve B try method A". The scientist's more hesitant dictum should take the form: "I had hypothesis X about the underlying mechanisms which control young children's development and this led to the prediction Y which I tested in this or that manner with some success". No one, I think, would dispute the value of either of these very different lines of debate, but the trouble is that the two sometimes become confusingly crossed. Bruno Bettelheim's new book is a case in point.

Bettelheim is a distinguished practitioner and a scientist too. For many years he has run a clinic for disturbed children. His ideas about childhood autism are well known and he has written interestingly and entertainingly about his work on children's reading and about his observations of the life of young children in Israeli kibbutzim. His latest book is about how

parents should behave towards their children.

The book's main themes are easy to describe. Bettelheim thinks that parents should reflect on the reasons for their own reactions towards their children, particularly when parent and child are in dispute. He argues that the real reason for these disputes often lies in the parent's own childhood difficulties and frustrations, and that once this is understood the parent will become more flexible and the difficulty should go away. Bettelheim also claims that children do not often do wrong knowingly and deliberately, and that parents ought to acknowledge this when they remonstrate with their child. Bettelheim abhors punishment and even warns parents against overwhelming a child with logic. Browbeating a child into admitting that he is in the wrong does no good and probably puts a child off rational argument into the bargain. In the final part of the book Bettelheim dwells on the importance of play and fantasy in childhood.

The book has many advantages. It is written in a relaxed and attractive way. Though many of the themes, such as the importance of play, are familiar ones, the anecdotes with which Bettelheim illustrates them are usually quite striking. It is also useful to be reminded of psychoanalytical accounts of relationships between children and their parents, because these have received little attention lately.

The trouble with the book is that it is full of confident assertions for which we have no real evidence. We can take as an example Bettelheim's claim that parents often unknowingly re-enact disputes which they themselves had as children with their own parents. The claim is plausible, and others have made much the same suggestion; Philip Larkin, for example, with his lines about parents: "They hand on all the faults they had, and add some extra just for you". Yet I know of no hard evidence for the suggestion that this happens, and worse still I could find no attempt in the book to argue against another possible explanation for why this pattern might occur. One alternative to the psychoanalytical explanation is genetic. Children may relive their parents' difficulties because they share their parents' genetic make-up.

Bettelheim's justification for his consistent failure to back his ambitious claims with anything but the flimsiest evidence might be that these are issues which are too complex for clear empirical research. But this excuse would not work with some of the issues in the book on which there has in fact been a great deal of research.

The most obvious example is the question of the effects of play. This has attracted the attention of a large number of research workers. It cannot be said that



Bruno Bettelheim — confident assertions.

their efforts have provided us with the final answer, but the evidence is certainly worth discussing and it ought to be mentioned by anyone making theoretical statements on the subject. Bettelheim's omission is a clear sign that he does not consider that a sound empirical base is needed for his theoretical ideas. That is a pity, because many of those ideas are interesting. □

P. E. Bryant is Watts Professor in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

The ANS Journals

YOUR
NUMBER ONE
SOURCE FOR CURRENT
INFORMATION ABOUT
**NUCLEAR
ENERGY**

#

- ☐ NUCLEAR NEWS
- ☐ NUCLEAR SCIENCE & ENGINEERING
- ☐ NUCLEAR TECHNOLOGY
- ☐ TRANSACTIONS
- ☐ ANS NEWS
- ☐ REMOTE SYSTEMS TECHNOLOGY
- ☐ FUSION TECHNOLOGY

FOR OUR CATALOG
CALL OR WRITE:

AMERICAN NUCLEAR SOCIETY
555 N. Kensington Avenue
La Grange Park, IL 60525 USA
312-352-6611

Poor old Charles Darwin

Antoni Hoffman

The Great Dying: Cosmic Catastrophe, Dinosaurs, and the Theory of Evolution. By Kenneth J. Hsü. *Harcourt Brace Jovanovich*: 1987. Pp.292. \$17.95, £15.60.

To make a long story short, the argument developed by Kenneth Hsü in his lucidly and entertainingly written popular book runs as follows: the iridium anomaly, osmium isotope ratio, shocked quartz grains and soot particle concentration in the Cretaceous/Tertiary boundary clay prove that a huge extraterrestrial object hit the Earth and caused mass extinction on land and in the sea. There is an ecological and temporal structure to these extinctions, and it demonstrates that the nuclear winter scenario does not fully explain the events. Some organisms were exterminated by a heat wave, others by subsequent darkness and freezing temperatures, still others by the surface ocean waters turning corrosive, or by acid rain, environmental pollution, or ultraviolet irradiation due to depletion of the ozone layer.

The ultimate culprit, nonetheless, was a catastrophe of cosmic dimensions. Thus the agent of species extinction was extrinsic to the biosphere instead of the competition from newly evolved and hence superior species. This conclusion falsifies the Darwinian theory of evolution through natural selection, because high fitness of a species does not guarantee its survival. Consequently, all the abhorrent ideologies demanding that stronger classes, races, nations or individuals take over and suppress or even exterminate weaker ones are deprived of their presumed 'scientific' basis and can be ultimately refuted.

Popular science writing has its *licentia poetica*, and a book review also has one. Hsü should no more and no less be blamed for his extreme one-sidedness than I should be for oversimplifying his argument and overreacting to his theses. But the reader should, I think, be forewarned that this book is one of unabashed partisanship. Hsü leaves absolutely no room for alternative explanations of the evidence. Yet science is all about choosing between alternatives, no matter how brilliant and exciting is the one a particular scientist espouses.

Hsü's partisanship may be deliberate,

• The Natural History Museum of Los Angeles County and University of Washington Press have just published Vol. 1 of *Dinosaurs Past and Present*, a large-format, well-illustrated book "for both the general reader and the specialist". The book is based on a travelling exhibition and a symposium, and costs \$35, £28.

intended to help spread the idea whose beauty has seduced him. Even such intentions, however, cannot justify the errors and misunderstandings in this book. For example, rudists did not flourish in the uppermost Cretaceous and disappear in the boundary clay; Digby McLaren's view of mass extinctions as geologically instantaneous catastrophes is not the consensus among paleontologists; Darwin wrote that interspecific competition leads to extinction, not that it is the only agent of extinction. Most importantly, the Darwinian theory does not claim that species which become extinct are those which have lower fitness, if only because fitness does not refer to comparisons between species. It is forgivable when a fiction writer like Kurt Vonnegut misses the meaning of fitness in an otherwise nice novel (*Galapagos*), but it is disappointing to see it happen in a book by an expert on mass extinctions whose explicit aim is to demolish and bury forever the "absurd", as he puts it, theory of natural selection.

Whether dinosaurs were actually killed by the impact of a comet, global climatic change, destruction of their habitats or hay fever triggered by angiosperm pollen, their extinction neither corroborates nor refutes Darwin's theory. And whether or not this theory is correct, it neither supports nor undermines the Nazi ideology and its like, which Hsü so rightly abhors. Whatever science may say about the laws of biological evolution — or for that matter about the nature of the physical world — it cannot possibly justify racism or Nazism, but it cannot disprove them either. Darwinism is no more responsible for the Holocaust and the mass killings of Poles and Russians as 'underpeople' than the Great French Revolution's call for liberty, equality and fraternity is to be blamed for the terror under Robespierre and Stalin. Hsü's partisanship has taken him too far. □

Antoni Hoffman is at the Institute of Paleobiology, Polish Academy of Sciences, 02-089 Warsaw, Poland.

Definitive doubts

J. F. Nagle

Phospholipid Bilayers: Physical Principles and Models. By George Cevc and Derek Marsh. *Wiley*: 1987. Pp. 442. \$79.95, £73.35.

THE appearance of several monographs on the physical chemistry of lipid bilayers is an indication that this field has come of age. The volume by D. M. Small (Plenum, 1986) that discussed the basic properties of many lipid systems is complemented by the more theoretically orientated book under review here, and both are complemented by B. L. Silver's well-written, more introductory account (Solomon Press, 1985).

In an introductory chapter, Cevc and Marsh review the results of many classical experimental techniques applied to bilayers. It is a little worrying that the brevity and the definitive tone of this chapter may lead readers to conclude that the experimental work is complete and all that remains is the theorizing typical of a mature or senescent field. In fact, many of the fundamental measurements have wide margins of error that create havoc when used in the kind of theory presented in the remainder of the book. An example is the enthalpy of transition of the best studied of all lipids, DPPC. Values in the recent literature range from 6.4 kcal mol⁻¹ to 9.3 kcal mol⁻¹, with the extreme values given by the two most experienced calorimetric groups and a scatter between provided by the rest of us. Naturally enough, in several cases where basic measurements are in dispute the authors have used their own

figures, but readers with alternative theories should be aware that some of the experimental values are not definitive.

After a fairly conventional chapter on the self-assembly of lipids into bilayers, there are several chapters dealing with the thorny topic of boundary water culminating in a chapter on interbilayer forces. The last half of the book reviews phase transitions and properties of the fluid phase, with much discussion of theoretical models and an effort to present some elementary statistical mechanical calculations. Each chapter contains a list of references which seems to be fairly representative up to 1984, but Cevc and Marsh often try to summarize the theoretical development of selected papers rather than simply presenting results. They also frequently present their own simple phenomenological treatments. Although the authors preface their book with the hope that their account will lead to further research, the overall tone tends to give the impression that the field is better understood than it actually appears to be to me.

Of the three books mentioned in this review, it is my opinion that the present one will be of the most immediate interest to researchers as a source of new ideas and ideas to argue with. Silver's book is rather more readable and would be my choice for a class or to introduce a graduate student to this field, while Small's large volume presents more data with a leaner superstructure of theory to hang them on. Certainly, each university library should not be without at least one of these books. □

J. F. Nagle is Professor of Physics and Biological Sciences at Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA.

NIH: a century of science for health

In 1887 a one-room Laboratory of Hygiene was established on Staten Island, New York, in response to public fears about the spread in the United States of yellow fever, cholera and tuberculosis. Direct descendant of that laboratory is the National Institutes of Health, based at Bethesda, Maryland, the largest and most successful organization ever dedicated to

biomedical research. At the end of a year of centennial celebrations, which culminated last week in a gathering of alumni on the campus at Bethesda, two former Directors, Robert Q. Marston and Donald S. Fredrickson, reflect on the salient events in NIH's recent history and look to the future of this giant of the biomedical research world.

Dilemmas of decision-making

Robert Q. Marston

ONE hundred years ago, the Laboratory of Hygiene on Staten Island was set up as a mission-orientated unit. That mission was health, and health has continued as and remains the object of NIH's endeavours.

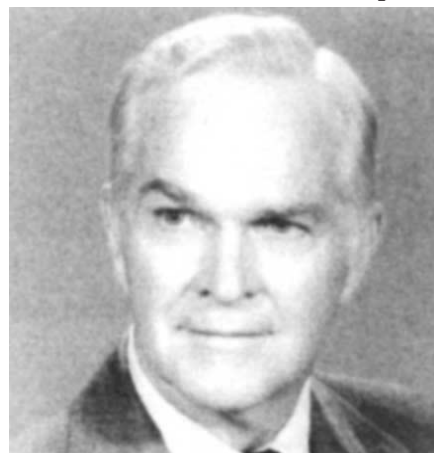
The modern NIH dates from the demobilization activities at the end of the Second World War, when it was given the responsibility for the administration of about 250 wartime research projects in progress at universities, medical schools and pharmaceutical companies. Thus began the powerful extramural programmes of NIH. These programmes, accounting for about 90 per cent of the NIH budget, permanently changed biomedical research through the sheer amount of money made available (almost five billion dollars), the number of institutions involved (1,650 of them, working on nearly 28,000 projects) and by the dependence on peer review of scientific merit as the crucial factor in the distribution of funds.

Since then, NIH has grown to include 12 separate research institutes and seven divisions or centres that support and promote research. It also includes the National Library of Medicine, the world's largest research library in a single scientific and professional field. Approximately 2,000 scientists conduct research in the laboratories at Bethesda. The activities in these laboratories and clinics, and the focus of constant meetings of study sections and councils debating the distribution of extramural funds, is on the use of science to improve health.

During the 1960s, great changes occurred in health care in the United States. Medicare and Medicaid were established as payment mechanisms for the aged and the poor. Through the Health Manpower Act, the federal government began direct support of education in the health fields. Because of its close affiliation with medical schools, NIH was given the

responsibility, for a few years, of implementing this legislation.

This was also a period of considerable emphasis on bringing the benefits of research to patients more rapidly. Success in the space programme had led many to believe that more effort, better organiz-



Robert Marston — "the success of biomedical research will be increasingly dependent on progress in understanding the basic nature of life".

ation and more money would accelerate the movement of research results from the laboratory bench to the patient's bedside.

Regional Medical Programs (RMPs) for heart, cancer and stroke were established in 1965 because of the belief in Congress and elsewhere that individuals and even institutions cannot cope with the complexities of modern medicine in isolation. Funds were provided to encourage and assist in the establishment of regional cooperative agreements among medical schools, research institutions and hospitals for research and study sections (including continuing education), and for related demonstrations of patient care in the fields of heart disease, cancer, stroke and associated diseases. Funds were also provided to afford to the medical pro-

fession and medical institutions, through such cooperative arrangements, the opportunity of making available to their patients the latest advances in the diagnosis and treatment of these diseases. For a few years, the NIH also had the role of implementing RMPs, and I came to NIH as the first Director of the Programs. Despite much excitement and considerable accomplishments in several parts of the country, RMPs passed away with many of the "Great Society" initiatives.

The fact that Health Manpower and Regional Medical Programs were placed in NIH underlines the special relationship between NIH and health educational schools throughout the United States. Over half of the extramural funds go to these schools. The advisory committees, both at the study section and the council level, are well represented by their faculties. For several years, the best of the young academics were able to fulfil their military obligations by a two-year period at NIH. Strong pre- and post-doctoral training programmes sponsored by NIH in the health educational institutions produced an additional bond.

By the early 1970s, NIH was back to the single responsibility for the conduct and support of biomedical research. Even its one regulatory component, the Bureau of Biological Standards, had been moved to the Food and Drug Administration. The drive to accelerate progress against dreaded diseases had not abated, and in 1971 the National Cancer Act became law. Among the many issues raised in this complex debate, none was more important to the scientific community than the question of targeted research versus investigator-initiated research. In short, the so-called "war on cancer" dramatized the continuing central question for all biomedical research institutions and research support institutions — how best to organize serious research, and who should make the decisions over priorities.

NIH has long supported a highly decentralized structure of semi-autonomous disease-orientated divisions, supplemented by several support units at the Director of NIH's level. The strong intramural programme has given administrators instant access to the substance of science; in addition, the Deputy Director for Science meets weekly with the scientific directors of the individual institutes. Hundreds of scientists convene in Bethesda as members of study sessions, councils and special advisory groups and add their expertise to the environment in which science policy is evolved. With this background, it is inevitable that science

policy, as viewed by the administrators of NIH, leans strongly towards the concerns of biomedical scientists rather than those of politicians and the general public.

The question of who should decide can be sharpened by quoting two sources that start at the opposite end of the debate. The first statement was made by Frank Carlucci when he was Undersecretary of Health, Education and Welfare (HEW):

There are two schools of thought . . . One says that a public agency should be a primary advocate of the special group of citizens that it serves. The other holds to the belief that a public agency must serve first and always the broad public interest, and take its direction and policy from the duly elected leader of the executive branch of government — be he president, governor, mayor or county supervisor. My beliefs, and those of Secretary Weinberger, rest clearly with this latter view. To me, public advocacy by a public agency is outright chaos. Sooner or later it places that agency in an adversary position with the chief executive.

This is in stark contrast to the attitude of the Minister of Health in Britain, speaking about the need to establish a Medical Research Council which was independent of government:

[A Minister of Health] would, therefore, be constantly tempted to endeavour in various ways to secure that the conclusions reached by organized research under any scientific body, such as the Medical Research Committee, which was substantially under his control, should not suggest that his administrative policy might require alteration . . . It is essential that such a situation should not be allowed to arise, for it is the first object of scientific research of all kinds to make new discoveries, and these discoveries are bound to correct the conclusions based upon the knowledge that was previously available and, therefore, in the long run to make it right to alter administrative policy . . . This can only be secured by making the connection between the administrative departments concerned, for example, with medicine and public health, and the research bodies whose work touches on the same subjects, as elastic as possible, and by refraining from putting scientific bodies in any way under

the direct control of Ministers responsible for the administration of health matters.

NIH has been helped over the years in handling this dilemma by an unusual and close working relationship with Congressional committees. In view of the size of its budget, it has been allowed a great deal of time to explain in detail what its priorities are and what its plans are, Institute by Institute. These testimonies before the Senate and the House Appropriation Committees are augmented by testimony from non-federal experts. The records of these testimonies then become the starting point of the next year's testimony. A similar process for detailed scientific discussions within the Executive Branch of government has never been fully developed, although from time to time the President's Science Advisor has helped fill this gap.

Although communication with Congress has been very good over the years, communications with the international scientific community are strengthened greatly by two NIH institutions. The intramural programme is the focus of the flow of colleagues from all over the country and the world, thus providing a remarkable opportunity for international scientific exchange. That communications system is augmented by another jewel at NIH, the National Library of Medicine. Having been established in 1836 as the Library of the Army Surgeon General's Office, the National Library of Medicine has just concluded its 150th anniversary celebrations this year. With materials in 70 languages and the capability of exchanging information internationally, the National Library of Medicine is today a worldwide link among all health professionals.

It has been difficult even in this year of celebration to quantify the contributions that NIH has made directly and indirectly. It has played a major role in the education of health professionals, in training young investigators, and in improving communications nationally and internationally

between biomedical researchers. A review of the scientific contributions made by those supported by NIH more than justifies its existence. Many believe that NIH has produced an effective, productive environment for discovery which is unique. As Lewis Thomas has said:

We are in the early stages of a genuine revolution in biological science that had its beginning in NIH starting around 40 years ago. All by itself, this magnificent institution stands as the most brilliant social invention of the 20th Century anywhere. It is also, by the way, an encouraging piece of evidence the government can once in a while do something exactly right and make it last.

Many factors have contributed to this success — a supportive Congress, decentralized management, dependence on universities for the conduct of most of the research, and a clear understanding of the different requirements for research and for development. Yet each of these factors has been under attack at one time or another. In the early 1960s, the Fountain Committee of the House of Representatives concluded NIH was not appropriately accountable to Congress; universities have worried about the unbalancing effects of the flow of funds into their medical schools; and there have been repeated efforts to change NIH into a more highly structured, centrally directed agency.

No other event in NIH's history displays so well the different views of how best to organize and manage biomedical research as did the National Cancer Act of 1971. Almost the entire biomedical scientific community, universities, medical professional organizations and NIH believed that the scientific base was lacking for a directed attack on cancer, feared a destructive breakup of NIH, saw a substitution of political judgement for scientific competence, and viewed with alarm the lure of quick success. By contrast, many cancer experts and the public became committed to a war on cancer. The administration first opposed and then



In the beginning — the Marine Hospital in Stapleton, Staten Island, birthplace of the one-room laboratory of Hygiene in 1887.

supported the National Cancer Act.

In his State of the Union message in January 1971, President Nixon recommended an "intensive campaign to find a cure for cancer", asked for an additional \$100 million for cancer research and stated: "I will ask later for whatever additional funds can be effectively used. The time has come when the same kind of concentrated effort that split the atom and put a man on the moon should be turned toward conquering this dread disease. Let us make a total national commitment to achieve this goal". At an earlier time, when the Nixon administration had tended to resist a proposal for a "war on cancer", I had been asked what I thought of such a programme, which, in addition to the possibility of diverting funds from other areas, proposed to remove the National Cancer Institute from NIH and to have it report directly to the President. I chose my first words with care. They were: "I am against it". This response was reported on the front page of the *Washington Post*. The fact was that, unlike the space programme and the atomic bomb, we simply did not have the fundamental knowledge to justify a centrally directed attack on cancer. Thus, I argued that greater faith should be placed on a more balanced investigator-generated grant mechanism than on a federally directed contract mechanism. There was no question, however, where the general public stood. This was dramatically demonstrated when Ann Landers's appeal in her column of May 1971 stimulated more than a million responses in support of the crusade against cancer. She thanked her readers with these comments:

The proposed National Cancer Authority would be an agency similar to the National Aeronautics and Space Administration which put the first man on the moon. If we stayed with the 'existing' machinery, it is doubtful that we would have reached that national goal. The answers to the scourge of cancer can and will be found. I say a massive unified assault on this killer disease is long overdue.

Fortunately, in the end, Congress did not vote on the basis of popular support alone. A compromise bill was worked out which retained the National Cancer Institute within NIH, but with a changed status. A similar status was established later for the National Heart Institute in recognition of the size and scope of its budget and responsibility. Funding for the Cancer Institute increased, passing the goal of a billion dollars per year, and significant progress has been made in the past 15 years—but nothing comparable to the hopes of those who wished to see the end of cancer in a decade.

It is of some interest that during these debates we were able to get a \$10 million increase in the budget of the National Institute of General Medical Sciences for the study of genetics. Although it is hard

The institutes

Forty years after the National Institute of Health formally became the National Institutes of Health, the number of institutes has grown to 12:

- National Cancer Institute
- National Eye Institute
- National Heart, Lung, and Blood Institute
- National Institute of Allergy and Infectious Diseases
- National Institute of Diabetes, Digestive and Kidney Diseases
- National Institute of Child Health and Human Development
- National Institute of Dental Research
- National Institute of Environmental Health Sciences
- National Institute of General Medical Sciences
- National Institute of Neurological and Communicative Disorders and Stroke
- National Institute on Aging
- National Institute of Arthritis, Musculoskeletal and Skin Diseases

The divisions and centres for the support or promotion of research are the John E. Fogarty International Center (for fostering international collaboration); the Division of Computer Research and Technology; the Clinical Center; the National Center for Nursing Research; the Divisions of Research Resources (a grant-awarding unit), Research Grants and Research Services; and the National Library of Medicine.

to prove, I have always thought that the rather timely infusions of modest funds was of considerable help in furthering the dramatic advances in this field.

I should make one brief comment about my role in the cancer crusade. I was very fortunate that my initial response about the Cancer Commission's recommendations was accurately and widely reported as "I'm against it". For some months, with the full support of the President's Science Advisor and the Secretary of HEW, I worked vigorously in Congress and elsewhere against the passage of the bill. Thus, by the time the administration had changed from opposing to supporting the bill, my position was entirely clear.

As one looks to the future, it is inevitable and appropriate that debate should continue on how best to support and organize biomedical research. Indeed, the fear of AIDS means that the same issues stimulated by the fear of cancer 17 years ago have risen to prominence once again.

NIH's contribution to progress in biological science as a whole has been great.

In the future, the success of biomedical research will be increasingly dependent on progress in understanding the basic nature of life. One effect will be the continued shift in the training of future NIH investigators. Clinical researchers already have difficulties in competing for grants unless they have extensive basic-science training or are a part of a team which involves basic scientists. Likewise, medical institutions without strong contacts with basic-science departments in universities are at a disadvantage which can only get worse.

Moral and ethical questions have required much more attention from NIH Directors than was needed in the past. There is every reason to expect this to continue in the future. I believe that we are still in the early stages of the revolution in our increased understanding of the biological world, and the future is quite unpredictable. The controlled experiment, the most powerful tool of our age, ensures that experiments come out as they come out, not as the investigator wishes. Crucially, future Directors must continue to be guided by the substance of science, wherever it might lead.

It would also be well to limit the responsibilities of NIH to the conduct and support of research. Even when its programmes were half the size they are now, diversion of efforts to health manpower and the organization of health care were carried out at some cost to the research activities.

Mr Carlucci's view of the need for federal agencies to be responsible to elected officials is valid. Indeed, Mr Carlucci is now back in Washington as President Reagan's National Security Advisor, because staff took it upon themselves to withhold information from elected officials. In that respect, it seems that the understanding of NIH, its role, capabilities and limitations, that was characteristic of the upper levels of the Executive Branch before and during my tenure as NIH Director, no longer obtains. According to numerous seasoned observers, that comprehensive, experienced capacity has been replaced by one that is primarily fiscal and political in nature. That situation is not in the long-term interest of either biomedical research or whatever administration is in power.

All research requires freedom from destructive political interferences, a high tolerance for eccentricity in creative individuals, and an environment that stimulates free and intensive intellectual activity. Biomedical researchers can justify even more protection as they work with the hopes and fears of all mortals in health, disease and finally in death. □

Robert Q. Marston is President Emeritus at the University of Florida, Gainesville, Florida 32606, USA. He was Associate Director of NIH from 1965 to 1968, and Director from 1968 to 1973.

Challenge of change in biology

Donald S. Fredrickson

REVISITING Bethesda for the Centennial, biomedical scientists from the four corners of the United States and abroad, including many who had postgraduate training and research experience at NIH, were able to check the pulse and reassure themselves of the continued vitality of their Alma Mater. In addition to being the world's largest single centre for biomedical research, NIH is also the disburser of about a third of the estimated \$15 billion annually spent on health research & development in the United States today. The NIH share also underwrites the bulk of basic research, and for decades has been the single largest source of support for academic research in the United States.

When the Laboratory of Hygiene moved to Washington in 1891, the Laboratory's small staff concentrated on infectious diseases and their control. They also discovered that pellagra was not infectious but a vitamin deficiency. This helped awaken legislators to the broader potential of federally conducted science, and in 1930 the Laboratory was converted into a National Institute of Health with expanded resources and objectives.

When Congress added a National Cancer Institute in 1937, it laid out future routes for taxpayer support of biomedical research. Extramural studies would also be supported. Diseases would be the prime targets, but the need for basic knowledge was understood and the search unlimited. The distribution of resources was delegated, not to institutions, but to the scientists themselves, whose decisions by peer review were overseen by citizen advisory councils. Thus 1987 is the fiftieth anniversary of NIH grants for investigator-initiated research.

Federal support of all science boomed at the end of the Second World War, and pluralization of NIH began in 1948 with a long procession of new organizational entities. The number reached 19 this year and includes the National Library of Medicine and its world-wide medical information system. Each has a separate appropriation, but shares intramural research facilities and administrative support under the direction of a Presidentially-appointed NIH Director.

Some thought it folly when the vast Clinical Center was opened in 1953. But the juxtaposition of 500 research beds and 1,000 laboratories was brilliantly successful in enhancing clinical investigation by directly linking it to basic research. In the period 1950–1965 physician scientists combining military duty and research at NIH became the vital source of faculty for expanding medical schools. As NIH

cloned itself, the numbers of researchers supported by extramural grants rose dramatically.

So did NIH appropriations, at first doubling about every three years. Early on, nearly 25 cents of every dollar went into training new scientists, a share that has fallen to a few pennies today. NIH also supported construction of new laboratories for many teaching institutions until the construction authorities ran out in around 1975.

NIH then had to turn to strengthening



Donald Fredrickson — “commercial biotechnology and academia have a necessary and beneficial symbiosis . . .”.

policies to deal with societal concerns over the rising power of biology. The Clinical Center procedures for clinical research were long preceded by those for animal studies. Both were eventually extended to all institutions where NIH funds research. NIH then became deeply involved in persuading lawmakers that voluntary guidelines were preferable to statutory control of revolutionary techniques for genetic recombination. NIH policies for improving intellectual property rights were extended to all federal research. Today the agency searches for consensus on medical interventions and exacts penalties for misconduct in science. Through all these demands, NIH remains an agency for research, not regulation.

Down the years, attempts also have been made to graft upon NIH other responsibilities sharply tangential to its research mission. These have included assurance of quality medical care to certain populations and the career development of health professionals or paramedical personnel. The Division of Regional Medical Programs and the Bureau of Health Manpower and Education, however, had but transient stays on the NIH campus. Because they can so easily haemorrhage away research funds, mandates for health care or control pro-

grammes have always been viewed with suspicion by NIH Directors. In the 1960s, the Director of the National Institute for Mental Health (NIMH) found irresistible the creation of an empire subsuming responsibility for centres for mental health care along with the research mission. The entire organization departed, leaving only the intramural laboratories at NIH. NIMH today is part of the Alcohol, Drug Abuse and Mental Health Administration (ADAMHA).

Like other sister agencies of NIH in the Public Health Service (PHS), ADAMHA thus contains programmes that began in NIH, such as early epidemiological activities, now in the Centers for Disease Control (Atlanta), or the Bureau of Biologics that directly descended from the old Hygienic Laboratory and is now part of the Food and Drug Administration. The shifting of these activities to other agencies with mandates for regulation or control has not only strengthened and better balanced the PHS, but spared NIH from dilution of its research functions. The intactness of the NIH coverage of human biology is marred today only by its separation from areas of neuroscience partially underwritten by the ADAMHA institutes. The threatened schism of the National Cancer Institute from NIH in 1971 — dealt with in the preceding article by Robert Marston — would have been a far more disastrous blow.

The paradox of NIH is that, while it is a public health agency, and part of the Department of Health and Human Services, it also had become a pre-eminent determinant of the strength of all academic medical centres and many research universities in the United States. This is because the NIH peer-review decisions have so much influence over which faculty members shall continue to do research and because the NIH grants and contracts include indirect costs of research — about a third of the total amounts — that uphold the infrastructure of the institutions.

The NIH budget is only a tiny fraction of the huge DHHS budget. The latter is overwhelmingly committed to fixed entitlement and benefit programmes, however, and the NIH budget now represents almost 40 per cent of the total discretionary funds. Thus considerable tension accompanies the annual Department budget and the Congressional appropriations for biomedical research.

For almost 40 years, from the beginning of the National Institute of Health, the Executive Branch and the Congress were united in pushing for medical research. In the past dozen years, however, the Congress has had to go it almost alone. The NIH appropriation has risen from \$2 billion in 1975 to over \$6 billion in 1987, overall an increase of about 28 per cent in purchasing power, when corrected for inflation, in costs of biomedical research.

In each of those years the budget of the Executive Branch has called for a cut in purchasing power from the previous year.

The Congressional support has been at least partially sustained by the adherence of both the House and the Senate to an agreement with the Administration, fostered by NIH in 1979–1980, to ensure that not fewer than 5,000 new and competing research grant proposals would be funded each year. Under this “stabilization initiative”, the total number of active research grants has risen from 12,000 in 1977 to 20,000 in 1987. Each appropriation bill has been accompanied by statements reaffirming the research project grant as the fundamental basis for federal support of biomedical research and a gradual increase in the minimum to over 6,000.

The stabilization formulae for awarding research project grants are of critical importance to an estimated 50,000 to 60,000 researchers supported by them each year. As successful as they have been in turning aside budget cuts, for a number of reasons the formulae will doubtless be put under further scrutiny in the future. First of all, many of the Institute leaders dislike the increasing pressure the quotas place on support for other extramural programmes, such as centres, training and clinical trials. Indeed, the portion of NIH extramural resources expended for these grants has risen from about 51 per cent to about 70 per cent over the past ten years. Quotas placed on grants also complicate experiments to increase the length of awards or to consolidate the funding of outstanding established scientists, which many research administrators and scientists agree is desirable. The average cost of a research project grant in 1987 was \$173,000. Because a top-notch scientist today may require more than a million dollars in annual support, adjustments in quotas would be required.

One of the unmentionable concerns for the future is the fact that NIH, from its inception, has managed to keep ahead of the pace of biomedical science by a slow but steady expansion. Were this to be prevented by some unavoidable manoeuvres associated with the federal deficit, there are no tested formulae for coping with a serious reduction. A mere “floor” for grants would probably not bear all the weight of revision required. Finally, the funding of biomedical research by so heavy an emphasis on small grants to individuals is under pressure from changes in the biosciences that are forcing many once autonomous scientists to become a part of a collective for survival.

NIH is in the vortex of such change itself. There are stresses that cannot be read in placid reflections of the first laboratories (built in 1940), or in the 12-storey glass surfaces of the adjoining Ambulatory Care Research Center



Laboratory complex — part of the National Institutes of Health 306-acre campus at Bethesda.

(1982). The science of life has been opened up to depths not imagined when NIH was born and the contours of biology and medicine are irreversibly changing. Molecular biology has also snatched away the ampersand that once comfortably separated health research & development. Already the cloning and sequencing of genes has become routine, and the greater passion is now to get the gene product, or change the gene and get an even better product. Scientists who have mastered the techniques of the “new biology” seek exciting uses for them, particularly in applications that affect health and disease.

As has already happened in many other disciplines, bioscientists and physicians now find themselves attracted to a relationship with industry, where the full continuum of research & development is maintained for profit. Commercial biotechnology and academia have a necessary and beneficial symbiosis and it seems that for most scientists the top premium placed on intellectual reward has not been devalued. Long excluded as government employees from benefiting financially from their work apart from their salaries, the intramural scientists at NIH have now been allowed to share in royalties on discoveries and inventions, and to consult (fees being limited to \$25,000 per year) or engage in collaborative research with industry. These associations with industrial biotechnology are here to stay, especially if the enormous public investment in the health mission of science is to be fully redeemed.

Part of intramural NIH has been moving into mission-orientated biotechnology for years. The National Cancer Institute (NCI) has received ample barbs in the past for its emphasis on targeted or directed research, while its more *laissez-faire* sibling Institutes brought Nobel prizes to Bethesda. Under Vincent T.

DeVita, the NCI now seems to be in high resonance with the times. First-class fundamental research is coupled with facilities for development and supplemented by contracts with industry; 149 of the invaluable beds in the Clinical Center are at the hub of a network of clinical trials. Such maintenance of the full continuum of biomedical research intramurally is part of the secret to keeping a Robert Gallo, co-discoverer of the AIDS virus, or a Steven Rosenberg, surgeon and pioneer in LAK cell therapy of cancer, at work in Bethesda.

And now comes AIDS. It may be the toughest test of NIH in this century. Congress and the Administration have cooperated in building appropriations for AIDS research to over \$900 million this year for NIH and the Centers for Disease Control. At the head of the research effort is Anthony Fauci, also Director of the National Institute of Allergy and Infectious Diseases. Scientific advances in AIDS research have come with unprecedented speed and new inter-agency cooperation is in place. There is concern, however, that the traditional loose coordination of efforts in government, academia and industry might not be enough for victory this time. NIH must be prepared to lead its powerful community into whatever configurations will get the job done.

Its work unfinished, NIH can look forward to at least another 100 years. There are many alumni who still see NIH as Camelot. So long as they remember that the enduring legends must change to conform with the times, they will be entitled to their illusions. □

Donald S. Fredrickson, National Institutes of Health, 6615 Bradley Boulevard, Bethesda, Maryland 20817, USA, was Director of Intramural Research in the National Heart Institute from 1968 to 1974, Director of NIH from 1975 to 1981, and President of the Howard Hughes Medical Institute from 1984 to 1987.

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50 – 80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963 – 65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

Neutrino spectroscopy of supernova 1987A

Lawrence M. Krauss

Departments of Physics and Astronomy, Gibbs Laboratory, Yale University, New Haven, Connecticut 06511, USA

Two of the four reported observations of neutrino bursts associated with supernova 1987A are consistent with each other and with the theoretical expectations of supernova neutrino luminosities. Testing a range of models for properties of the incident neutrino signal, together with calculation of interaction rates in the chlorine solar neutrino detector, helps to provide constraints on the mass and number of light neutrinos and raises the possibility of vacuum neutrino oscillations.

ANY observation of neutrinos emitted by a supernova has great significance for both particle physics and astrophysics. It would be the first direct glimpse of stellar core collapse and neutron star formation inside a supernova¹⁻³, and would provide a fundamental test of the theory of stellar evolution. The characteristics of the neutrino signal may also give information on the weak interactions and on the physics of neutrinos and perhaps exotic particles.

Four different large underground detectors reported⁴⁻⁷ observations of neutrino bursts on 23 February 1987, all within 24 hours of the visual sighting. It is important to determine if any combination of these observations can consistently be associated with a single model for the expected supernova neutrino signal. Credible data from more than one group is very useful if reliable modelling is to be done, especially because of the small number of events in each experiment. The statistical significance of a putative feature in one set of data can be examined by comparing with similar data from another experiment.

I have performed a detailed analysis of the neutrino signals from the Kamiokanda⁶ and IMB⁷ water detectors, the expected signal in the chlorine detector of Davis *et al.*⁸, and to a lesser extent the signal in the Mont Blanc detector⁴. The analysis incorporates all the relevant neutrino interaction processes and the reported resolution functions (M. Goldhaber, personal communication) of the detectors. Combining this with model input I performed χ^2 and Monte Carlo likelihood analyses of the data to examine for internal consistency between the different experiments, and to determine the confidence with which assertions about different models may be made. The purpose of carrying out so detailed an analysis on such sparse data is not to provide definitive evidence in favour of one specific model, but rather to ascertain how wide a range of parameters may fit the data with any confidence, to investigate any apparent discrepancies between the different observations, and to determine the significance of various features of the signals. Nevertheless, the data do favour certain models and appear inconsistent with others. In addition this analysis allows more insight into what, if any, novel neutrino physics may actually be gleaned from the data. The general conclusions are these: (1) The Kamiokanda and IMB data agree well with theoretical expectations for the neutrino luminosity from a supernova ($1.5\text{--}3 \times 10^{53}$ ergs), even if some of the burst events observed in each detector are due to noise. Spectral analysis gives a best fit of ~ 4.5 MeV for the mean temperature of $\bar{\nu}_e$ emission from the supernova core although a wide range of different temperature-luminosity combinations is consistent with the data. (The Mt Blanc observation is, however, extremely difficult to reconcile with either of these experiments or with supernova theory.) (2) The timing of both observations is consistent with an initial luminosity decreasing as an exponential or gaussian with a time constant of > 1 s. Both the spectral and timing information suggest that at least some of the late events are likely to be noise-related, or that there is additional late-time structure in the emission. In particular, a

lower emission temperature seems to be required for the late events (> 6 s), indicative of cooling. (3) Conservative neutrino mass limits derivable from the timing of the neutrino bursts are probably weaker than one might initially expect. The limit on the number of neutrino species is also weak ($N < 9$). Finally, the data are quite consistent with vacuum neutrino oscillations at a level that could resolve the solar neutrino problem.

Model analysis

Cross-sections. The neutrino interactions that can deposit enough energy to be measurable as Cerenkov light are ν (or $\bar{\nu}$)-e scattering and $\bar{\nu}$ -induced p-n conversion and nuclear transmutation of oxygen. Cross-sections used in the analysis for these processes are shown in Table 1.

The dominant process is $\bar{\nu}$ -p scattering. Because of the large target mass, at the energies of interest almost all the neutrino

Table 1 Relevant neutrino cross-sections, and energy deposits

$\nu(\bar{\nu})$ -e scattering	
$d\sigma/dT = G_F^2 m_e / 2\pi [(g_v + g_a)^2 + (g_v - g_a)^2 (1 - T/E)^2 - (g_v^2 - g_a^2) m_e T/E^2]$	
$T_{\max}(e) = 2E^2 / (2E + m_e) v_e(\bar{\nu}_e)$; $g_v = 2 \sin^2 \theta + 1/2$; $g_a = 1/2(-1/2)$	
$v_{\mu,\tau}(\bar{\nu}_{\mu,\tau})$; $g_v = 2 \sin^2 \theta - 1/2$; $g_a = 1/2(-1/2)$	
$\sigma_{\text{tot}}(\nu_e) \approx 8.7 \times 10^{-44} (E/10 \text{ MeV}) \text{ cm}^2$	
$\bar{\nu}$ -p scattering	
$\sigma_{\text{tot}} \approx G_F^2 (C_V^2 + 3C_A^2) m_p E^2 / \pi (m_p + 2E) [1 - (m_n^2 - m_p^2) / 2m_p E]^2$	
$\bar{T}(e^+) \approx E - E^2 / (m_p + 2E) [1 - (m_n - m_p) / E] - 1.29 \text{ MeV}$; $C_V = 1$;	
$C_A = -1.25$	
$\sigma_{\text{tot}}(\bar{\nu}_e) \approx 7.5 \times 10^{-42} (E/10 \text{ MeV})^2 \text{ cm}^2$	
$\bar{\nu}$ - ^{16}O scattering	
$\sigma_{\text{tot}} \approx (2I' + 1) / (2I + 1) [G_F^2 m_p (E - E_{\text{thr}}) p_e / ft_{1/2}]$	
$T(e^+) \approx E - E_{\text{thr}}$	
$\sigma_{\text{tot}}(\bar{\nu}_e) \approx 9 \times 10^{-42} ((E - E_{\text{thr}}) / 10 \text{ MeV})^2 / (1,000 / ft_{1/2}) \text{ cm}^2$	

energy is transferred to the outgoing positron, so only the total cross-section need be considered here. We have included a factor that takes into account the mean energy transfer to the nucleus both in determining the mean energy ($\bar{T}$) of the outgoing positron and also in fixing the overall magnitude of the cross-section. The differential cross-section is nearly flat in the kinematic range of interest, but owing to dominance of the $(C_V - C_A)^2 (1 - T_{\text{nuc}}/E)^2$ piece there is a small (10–20%) enhancement in the forward direction for higher neutrino energies.

The ν -e scattering cross-section is dominated by electron neutrinos. Although the differential cross-section $d\sigma/dT$ is flat, the scattered electrons are preferentially directed forward. The overall cross-section, which rises linearly with energy, is over 70 times smaller than the $\bar{\nu}$ -p cross-section for energies > 10 MeV.

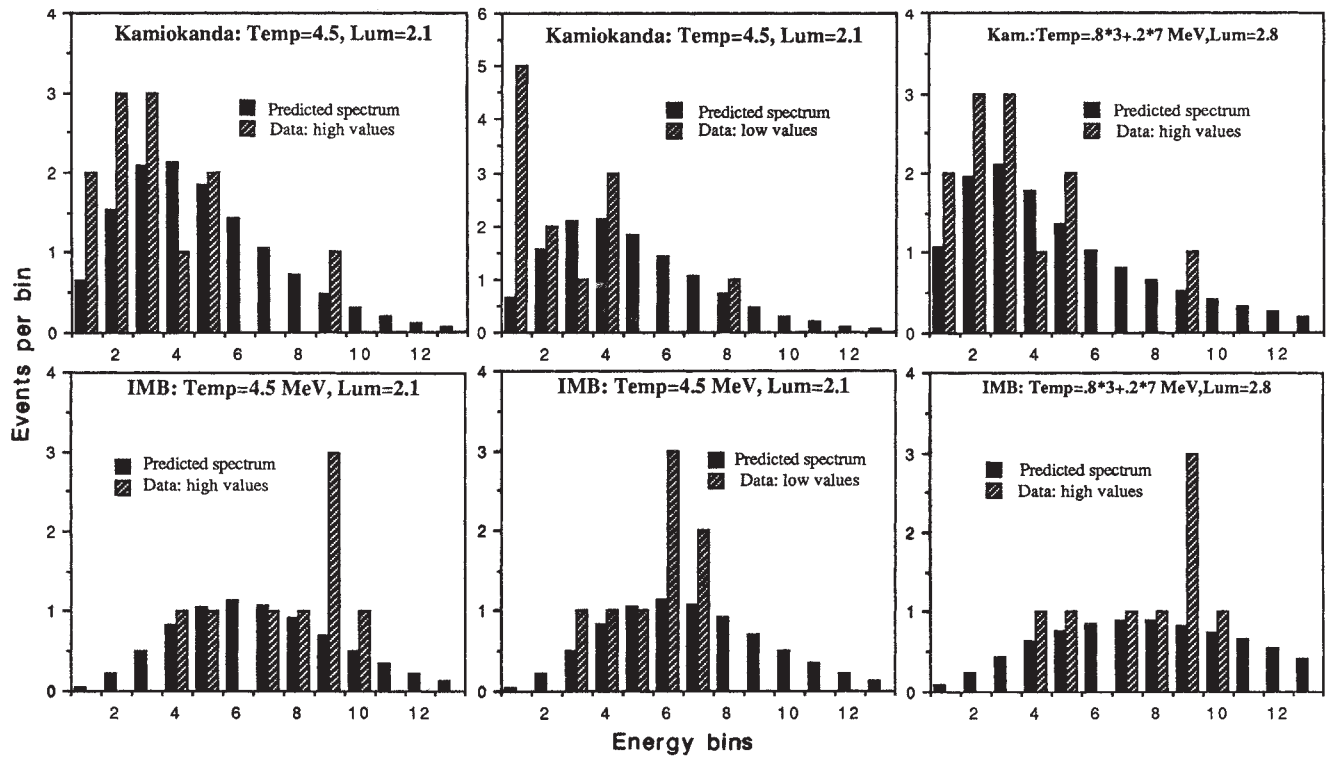


Fig. 1 The observed signal in both the Kamiokanda and IMB detectors, along with time-integrated binned spectra from a supernova neutrino burst with a thermal $\bar{\nu}_e$ spectrum of $T = 4.5$ MeV, and a net luminosity in all neutrino types of 2.1×10^{53} ergs, as well as a spectrum arising from a linear combination of two thermal distributions (see text). The 13 energy bins range from 5.7 to 54.3 MeV. (The upper endpoints of each bin are 9.6, 13.5, 17.4, 21.2, 25.2, 28.8, 32.5, 36.3, 39.9, 43.6, 47.2, 50.8, 54.3 MeV.) 'High' implies systematic use of the upper end of the measured energy ranges, and 'low', the low end.

Antineutrino neutrinos can induce the reaction $\bar{\nu} + {}^{16}\text{O} \rightarrow e^+ + {}^{16}\text{N}$. (${}^{17}\text{O}$ and ${}^{18}\text{O}$ are so rare that we ignore their induced transitions to N or F at these energies.) This has a threshold of ~ 10 MeV. The ground-state transition has a large $\log ft_{1/2}$ value of 6.7. Any transition to an excited state is expected at best to be allowed but unfavoured, with $\log ft_{1/2} \approx 4$. For purposes of calculation we have assumed such a transition, with a threshold of ~ 13 MeV, where several candidate states lie. The resulting cross-section is still generally less than 5% of the $\bar{\nu}$ -p figure.

Spectral fits. The above cross-sections were numerically integrated over an incident thermal neutrino spectrum (described below), a fitting function which parametrized the triggering efficiency of the Kamiokanda and IMB detectors^{6,7}, and a gaussian energy resolution function. This yielded a prediction for the spectrum of energy deposited in the detectors. Both the predicted spectra and the data were then binned before they were compared using a χ^2 analysis.

The number density for each neutrino type was assumed to have a Fermi-Dirac form, $n(E) dE = NT^{-4} E^2 dE / (\exp(E/T) + 1)$, where T is the emission temperature. Theory suggests¹⁻³ that ν_e s can have a lower emission temperature (due to their higher opacity) and this possibility was allowed for in the analysis. The overall normalization N was obtained by setting the integrated spectral density ($\propto T^4$) for the six ν helicity states to be equal to N times 10^{53} erg. The number density of each species was then taken to be $N/6$ times the standard Fermi-Dirac distribution function. (In fact, a somewhat greater fraction of ν_e s is expected¹⁻³. Although this will not affect the calculated rates, because they are primarily due to interactions of $\bar{\nu}_e$ s, it will imply that the calculated luminosities should probably be considered lower bounds.) The flux was finally calculated assuming a distance to the Large Magellanic Cloud (LMC) of 50 kpc.

The observed distribution of events depends strongly on the triggering efficiencies of the detectors. Using the quoted values^{6,7}

I defined a function, $F = 1 - \exp(-(E/E_{\text{thr}})^p)$ to mimic the actual efficiencies, with $E_{\text{thr}} = 34$ MeV and $p = 3.1$ for IMB, and $E_{\text{thr}} = 9$ MeV and $p = 3$ for Kamiokanda. The final event rates were then determined assuming fiducial masses of 2,140 and 5,000 tonnes for Kamiokanda and IMB respectively.

An energy resolution function of gaussian form with width $\approx (E_{\text{observed}})^{1/2}$ (in rough agreement with the reported widths) was

Table 2 Predicted rates and χ^2 values for various model spectra

Detector	Temperature (s)* (MeV)	Luminosity (10^{53} erg)	Predicted number of events	χ^2 value
Kamiokanda	3	4	12.9	11.3
IMB	3	4	3.7	84.4
Kamiokanda	4	2.5	12.8	7.2
IMB	4	2.5	6.4	17.5
Kamiokanda	4.5	2.1	12.6	9.5
IMB	4.5	2.1	7.7	10.5
Kamiokanda	5	1.8	12.3	13.4
IMB	5	1.8	8.9	7.7
Kamiokanda	6	1.5	12.4	23.9
IMB	6	1.5	11.7	6.5
Kamiokanda	7	1.3	12.2	39.1
IMB	7	1.3	13.8	7.4
Kamiokanda	$0.85 \times 4 + 0.15 \times 8$	2.1	12.3	9.6
IMB	$0.85 \times 4 + 0.15 \times 8$	2.1	8.6	8.7
Kamiokanda	$0.8 \times 3 + 0.2 \times 7$	2.8	12.5	6.5
IMB	$0.8 \times 3 + 0.2 \times 7$	2.8	8.0	9.2
Kamiokanda	$0.7 \times 4 + 0.3 \times 7$	2	12.8	11.5
IMB	$0.7 \times 4 + 0.3 \times 7$	2	9.9	7.2

* Temperature is $\bar{\nu}_e$ emission temperature. It is assumed here for the purpose of quoting luminosities that the $\bar{\nu}_e$ flux is 1/6 of the total flux. χ^2 values for Kamiokanda data are calculated using the high end of allowed range for quoted energies.

included in the integrations. Binning was carried out with bin size roughly twice the smallest resolution reported, or ~ 4 MeV. This was coarse enough that a χ^2 analysis could give reasonable preliminary evidence for the consistency of the data with various hypotheses of overall temperature and luminosity. The predicted spectrum was evaluated for temperature T and luminosity N , then divided into 13 energy bins, from ~ 5.7 to 54 MeV. Because the mean energy imparted to the O nucleus increases with increasing ν energy, a uniform distribution of neutrinos will yield an e^+ energy distribution which is not quite uniform. This effect was taken into account, so that the highest-energy bins turn out to be 0.4 MeV narrower than the lowest-energy bins. The actual data was then placed in similar bins, allowing comparison with models.

Timing and likelihood analysis. The complete time-dependent signal was then analysed to decide, for example, whether one or several bursts is more consistent with the data, what time constant best fits the data, and whether the temperature was constant over the period of the burst(s). In addition, a Monte Carlo statistical likelihood analysis was performed so more precise statements of confidence in different models could be made. A time grid of separate half-second intervals spanning a 13-s period around both sets of observations was created. For each time slice, 13 separate energy bins were used, corresponding to the same spectral binning done earlier for the time-integrated signal. The various predicted model distributions, as well as the data, were then stored in 26×13 arrays for later use in the likelihood analysis. The choice of such a time grid was not crucial to the analysis. As will be shown below, this binning procedure yields likelihood results essentially identical to those obtained for a more conventional unbinned sample, as long as Poisson statistics are used.

The different luminosity functions used all took the form

$$L = A/\tau(2/\pi^{1/2})^{n-1} \exp(-(t/\tau)^n) + B/13 \\ + C(1/\pi^{1/2}) \exp(-(t-11\text{ s})^2),$$

where $n=0$ or 1 so that the variables A , B and C represent in combination absolute luminosities in units of $10^{53} \text{ erg s}^{-1}$. The first term represents an exponentially falling or gaussian initial pulse with time constant τ , the second term a possible constant luminosity factor, and the third term a possible delayed burst resulting in the last three observed Kamiokanda events. For each of these luminosity functions three different temperature distributions were considered: a constant temperature, a fast increase followed by a slow decrease ($T/\text{MeV} = 3 + 2t(t < 2\text{ s})$, $T/\text{MeV} = 7 - 0.4t(t > 2\text{ s})$), and a stepwise decrease ($T = \text{const}(t < 6\text{ s})$, $T = \text{const}/2.4(t > 6\text{ s})$).

Let the reader be concerned about the number of variables used, I reiterate that the parametrization above was created to allow a wide variety of different distributions to be compared with the data. Best-fit modelling was not attempted for all these parameters. Their purpose was rather to allow an examination of exactly how varied were the distributions which might be consistent with the data.

Two different likelihood analyses of the models were carried out. First, the 13 seconds in question were separated into 0.1-s intervals. In each interval the expected number of events in each energy bin was used to define a Poisson weighting function for a Monte Carlo procedure in which a candidate number of events were randomly assigned to each bin. The results were amalgamated into 26 time slices for each of the 13 energy bins, to match the binning of the actual data. In this way many different candidates for observed distributions were generated. For each such distribution, Poisson likelihood values were determined by taking the logarithm of the probability that those distributions that contained the observed number of events might result from the theoretical postulate. The percentage of candidate distributions (expected to be accurate to $\leq 9\%$) with a likelihood value less than that of the observed distribution, multiplied by the

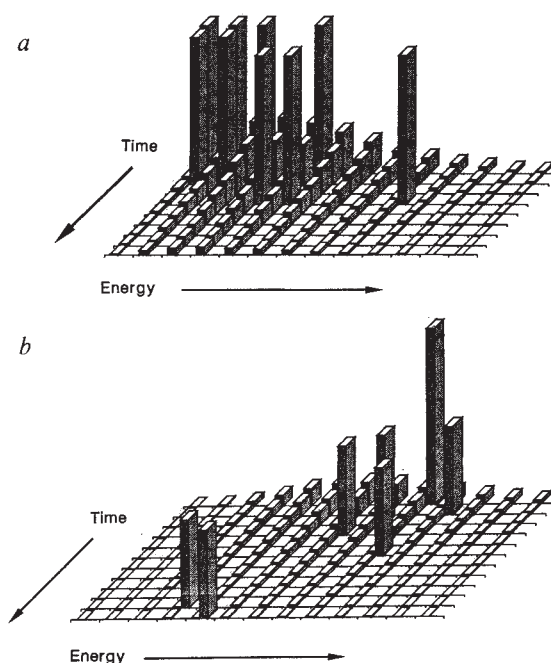


Fig. 2 *a*, The first 4.5 s of the Kamiokanda data compared to the predicted distribution (in 0.5 s time bins and 13 energy bins) for a 5-MeV neutrino burst with an exponential fall-off characterized by a 2-s time constant. *b*, The first 6 s of the IMB data compared to the predicted distribution for the same parameters.

probability that the observed number of events would result from the predicted distribution, divided by the same probability for a distribution with a mean number equal to the observed number of events, was then used as a confidence indicator.

The large number of bins (26×13) compared to the small number of events might cause concern. But because each bin effectively contains either zero or one event, the Poisson likelihood function becomes $P = \sum \ln \mu_n + \text{const}$, where μ_n is the expected number of events per bin and the sum is over non-zero bins only. The large number of zero bins merely yields a constant equal to the total number of expected events. It is then straightforward to show that in this limit the likelihood analysis is stationary as the number of bins is increased, and yields the continuum time result.

Finally, the percentage of candidate distributions with the observed number of events and having no signal for a period of 7 seconds or more in either detector (which occurs strikingly for the observed data) was also recorded as a likelihood indicator. A variable delay between the first Kamiokanda event and the first IMB event was also allowed to check whether such a possibility might better fit the data.

Predictions for chlorine detector. The expected event rate in the chlorine solar neutrino detector⁸ was also calculated for the different models in the above analysis. The numerical algorithm used had been developed earlier to calculate the response of this detector to the solar neutrino spectrum, and has an expected error of $< 10\%$ in the calculated neutrino capture rate⁹.

Estimates of Mt Blanc signal. For the purposes of approximate comparison between the Mt Blanc signal and the expected signals in Kamiokanda and IMB, I treated the former detector, which is a liquid scintillator detector, by the same scheme used to predict rates in the other water detectors. The resolution function was similar in shape to that used for Kamiokanda, but with $E_{\text{thr}} = 5$ MeV and fiducial mass ≈ 100 tonnes. Assuming that $\bar{\nu}_e$ -p scattering dominates in all detectors, this should give a reasonable idea of the expected event distributions.

Table 3 Monte Carlo results for various models

Model*	Luminosity decay time	Temperature (MeV)	Luminosity (10^{53} erg)	Predicted events		Confidence (%) [†]		% (>7-s gap) [‡]
	(s)			Kam.	IMB	high	low	
<i>Exponential decay</i>								
A	3	5	1.9	12.6	9.2	58	38	5
A	2	5	1.8	12.0	8.8	14	7	30
B	2	5	1.9	12.3	8.8	4	3	66
B	3	5	1.9	11.8	8.2	40	20	26
B	5	5	2.2	12.0	8.1	78	59	4
B	∞ [§]	5	2.4	11.7	6.7	85	69	0
<i>Exponential decay plus constant-luminosity tail</i>								
A	3	3	4.2	13.2	3.8	7	15	1
A	3	4	2.7	13.5	6.7	80	80	0
A	3	4.5	2.1	12.3	7.6	92	88	2
A	3	5	1.8	12.0	8.7	92	76	4
B	3	3	4.2	11.4	3.1	1	5	9
B	3	4	2.9	11.8	5.5	30	46	19
B	3	4.5	2.7	12.6	7.3	62	65	25
B	3	5	2.1	12.2	8.2	88	69	9
B	3	6	1.5	10.6	9.4	47	12	15
B	3	$0.75 \times 4 + 0.25 \times 8$ [¶]	2.4	12.3	9.3	85	66	10
B	2	4.5	2.5	12.7	7.5	57	63	33
B	2	5	2.1	12.6	8.6	75	57	16
B	2	$0.75 \times 4 + 0.25 \times 8$	2.4	12.8	9.8	63	48	26
B	4	4.5	2.7	11.9	6.8	56	63	15
B	4	5	2.1	12.0	7.9	94	75	2
B	4	$0.75 \times 4 + 0.25 \times 8$	2.3	11.5	8.5	91	71	5
B	1	5	1.8	11.0	7.8	25	19	46
<i>Exponential decay plus second burst</i> [#]								
B	2	5	2.2	12.3	8.0	74	57	11
B	2	$0.75 \times 4 + 0.25 \times 8$	2.5	12.0	8.8	86	68	21
B	3	5	2.4	12.5	7.7	95	90	3
B	3	$0.75 \times 4 + 0.24 \times 8$	2.6	12.0	8.6	92	83	10
<i>Exponential decay; last three Kamiokanda and one IMB events removed</i>								
	2	5	1.5	9.7	7.0	95	91	70
	2	$0.75 \times 4 + 0.25 \times 8$	1.5	9.0	7.2	98	92	71
	3	5	1.5	9.3	6.5	98	98	39
	3	$0.75 \times 4 + 0.25 \times 8$	1.7	9.4	7.3	98	93	45
	4	5	1.6	9.3	6.4	97	96	20
	4	$0.75 \times 4 + 0.25 \times 8$	1.7	9.0	6.9	100	100	25

* A means temperature was constant; B means the temperature at times after 6 s was $\sim 1/2$ the temperature before (rounded up to nearest integer).

† Percentage of Monte Carlo distributions with Poisson likelihood value less than observed distribution, multiplied by probability that model distributions lead to the observed number of events in each detector. 'High' implies systematic use of the upper end of the measured energy range; 'low', the low end.

‡ Percentage of simulations which had a 7-s or longer interval with no events.

§ Constant luminosity.

|| For 'A' models, luminosity in constant piece was $\sim 30\%$ of that in exponential, for 'B' models it was 40–50%.

¶ Mixed populations of electron and muon neutrinos at lower and higher temperatures.

Gaussian population centred on 10.5 s.

Results

Table 2 shows the integrated event rates for different temperatures and luminosities, along with the χ^2 values. In Fig. 1 several of these integrated spectra are shown along with the data. (The temperature shown is the $\bar{\nu}_e$ temperature. As discussed below, the results were insensitive to the temperature of the other neutrino distributions, which was taken to be ~ 8 MeV.) These features are examined in greater quantitative detail by the timing and Monte Carlo analyses. Fig. 2 shows the temporally binned spectra for one choice of model parameters; Table 3 shows the results of the Monte Carlo likelihood analysis for several widely varying models. Several conclusions can be drawn.

The relative and absolute rates in the two detectors agree well, for $\bar{\nu}_e$ temperatures less than 7 MeV, on the assumption of equal luminosity in each neutrino species and total neutrino luminosities in the range $(1-4) \times 10^{53}$ erg, in good agreement with standard theoretical models¹⁻³ of neutrino emission from a single supernova at the distance of the LMC.

The expected number of events induced by neutrinos other

than $\bar{\nu}_e$ s is less than one. As a result the predicted spectra are very insensitive to the initial ν_μ and ν_τ temperatures.

For a single $\bar{\nu}_e$ thermal distribution, a best fit to both sets of data has the temperature in the range 4–5 MeV, with Kamiokanda favouring the lower and IMB the higher value. At a 90% confidence level, the range 3–7 MeV is acceptable. (For a Fermi distribution, the mean energy $\langle E \rangle = 3.53 T$. This implies a marginally higher mean energy than some theoretical estimates¹⁻³.) It is worth remarking how good are some of the fits to the data within this range.

The Kamiokanda data agree marginally better with the models and with IMB if the high end of the allowed energy range for each data point is chosen.

Somewhat better agreement is also possible between both signals if some of the later, low-energy Kamiokanda events are due to noise. The best-fit temperature then rises to be closer to 5 MeV.

If one chooses *a priori* a $\bar{\nu}_e$ distribution that might be expected to result from vacuum neutrino oscillations (and not to fit the data), characterized by a linear combination of thermal distribu-

Table 4 Interaction rates in chlorine detector for supernova signal

ν_e Temperature (MeV)	Total luminosity (10^{53} erg)	ν_e flux* ($\text{cm}^{-2} \text{s}^{-1}$)	Interaction rate (SNU s = number per 10^{36} atoms)	Event rate (610 tonnes)
2	2.5	1.4×10^{10}	1.6×10^4	0.04 events
3	2	7.4×10^9	3.3×10^4	0.07 events
4	1.8	5.0×10^9	5.1×10^4	0.11 events
5	1.6	3.6×10^9	6.5×10^4	0.14 events
7	1.4	2.2×10^9	8.6×10^4	0.19 events

* Assumes ν_e flux = 1/6 total.

tions with two different temperatures, the fit can improve. While one can always obtain a better fit by adding an extra parameter, this new distribution was chosen for theoretical reasons unrelated to any best-fit analysis. In this case there is no guarantee that the fit will improve, and it is therefore interesting that it does for such a distribution, which is broader than a thermal one.

Temperatures below 3 MeV would allow a burst of three or fewer events in the Kamiokanda detector, with no events observed at IMB. This might be relevant to some or all of the late events at Kamiokanda ($t=9-12$ s).

A strongly varying temperature does not fit the data for the first six seconds better than the constant- T models described above.

As far as the time structure of the neutrino signal is concerned, these analyses alone would suggest that a wide range of parameters are compatible with the observations, ranging from a constant temperature and constant luminosity to an exponentially falling luminosity with a time constant >2 s. The model that seems to fit best, but by no means definitively so, is an exponentially falling initial signal with a time constant of the order of 2–4 s, in which at least 3/4 of the total neutrino luminosity is contained. The Kamiokanda data favours a shorter time constant than the IMB data. It can also be seen that better fits are possible if this initial signal is followed by additional structure. In this case, a shorter (~ 1 s) time constant for the initial burst can be consistent with the data.

The Poisson analysis is, however, fairly insensitive to longer time-correlations of the signal. As long as the expected number of events per bin is small and roughly constant, the same Poisson likelihood value will result for events that are equally spaced in time, or bunched into a few sets of adjoining bins. Simply scanning by eye the Kamiokanda data, one is struck by the apparent time structure of the signal. Most notable is the 7-s interval between the initial nine and the final three events. One senses that this is improbable for a simple exponentially falling luminosity function, but given the different sensitivities of the two detectors it is possible that additional late emission at low temperature can be reconciled with the 7-s interval at IMB where no events are seen? The second likelihood estimator, sensitive to the number of consecutive zero bins, gives a better way to probe this time structure for the various models of temperature and luminosity.

With this additional piece of information included, the possibility that the some of the three late Kamiokanda events are due to noise appears more likely. A 7-s interval with no events is not very probable, but a wide range of possibilities remains consistent with all the data. The models that best accommodate the late time signal all involve lower temperatures at late times, as one might expect for a cooling model. For example, a low-temperature exponential or roughly constant tail, or a second smaller, late, low-temperature burst are possible. Including this additional timing information also constrains the time constant of the initial burst to be less than 5 s. The timing agreement between Kamiokanda and IMB is improved if the last IMB event is also due to noise.

Adding noise has other implications. If the last three Kamiokanda events (and perhaps the last IMB event) are not due to the supernova, then the best-fit temperature is increased

slightly. In addition, the mild anisotropy of the observed Kamiokanda events is further increased compared to expectations, but not to the extent that one would require all the events to be related to the supernova.

The predicted event rate in the chlorine solar neutrino detector, given in SNU s (where SNU is solar neutrino unit) and in terms of the total number of events, is given in Table 4, where it can be seen that although the different models predict very different interaction rates, the integrated rates are all expected to result in fewer than one event.

Finally, the result of fitting the Mt Blanc burst (which occurred ~ 5 h before the burst discussed above) to a thermal spectrum is shown in Fig. 3. A good fit is obtained by a 1-MeV burst with a luminosity of 500×10^{53} erg! In this case many events would be expected in the Kamiokanda detector, and some at IMB. Of course, even if the signal were supernova-related, it need not be thermal, and a yet lower temperature could reduce the expected number of events in the two water detectors to zero, but the required luminosities are indicative of the problems of associating the Mt Blanc observation with new supernova physics. Even ignoring any attempt to fit to the data, the roughly 10 to 1 probability of seeing an event in the energy range of 5–10 MeV in Kamiokanda as opposed to Mt Blanc suggests that there is an *a priori* probability of $<1\%$ of seeing five events at Mt Blanc and seeing less than two events at Kamiokanda, which further demonstrates the difficulty of ruling out other sources for these events.

Conclusions

The analysis described here suggests that on the basis of spectral shape and overall event rate, both the Kamiokanda and IMB observations of neutrino bursts on 23 February 1987 are consistent with a supernova source in the LMC. Based on this initial result, conclusions can be drawn of relevance to both particle physics and astrophysics (see also refs 12–15).

The standard theory of supernova dynamics agrees remarkably well with the observed neutrino luminosity from supernova 1987A, assuming that roughly equal numbers of neutrinos and antineutrinos of each type are emitted with luminosities in the range $(1.5-3) \times 10^{53}$ erg, although a somewhat wider range remains consistent with the data. The lower bound is increased slightly if the predicted excess of ν_e s is allowed for. A null result from the chlorine solar neutrino detector will limit the electron neutrino luminosity to be ≤ 10 times greater than the highest value in this range. Thus, neutrinos carry off almost all the gravitational binding energy released as the supernova core collapses to form a neutron star, and a significant fraction of the supernova neutrino luminosity is carried by antineutrinos.

If a single emission temperature is assumed, then the predicted mean energy of electron neutrinos emitted from the supernova is best fit by a temperature of ~ 4.5 MeV, marginally higher than some estimates but in good agreement with theory¹⁻³. The mean neutrino energy is then slightly more than 15 MeV. A wide range of temperatures, 3–7 MeV, are nevertheless consistent with the data. These results verify the expectation¹⁻³ that the dominant supernova neutrino cooling process results from thermalized neutrino pairs diffusing from the edge of the neutrinosphere, and not from deleptonization processes happening inside the

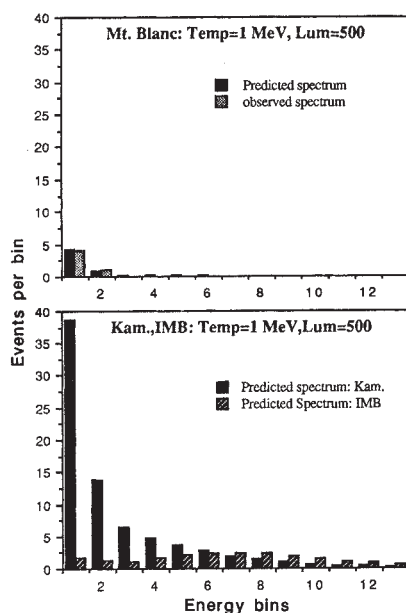


Fig. 3 Using the same algorithm used to calculate the predicted spectrum in IMB and Kamiokanda, modified to take into account the different threshold and volume of the Mt Blanc detector (see text), the predicted spectrum in all three detectors is shown for a thermal neutrino burst with temperature 1 MeV, and luminosity 500×10^{53} erg.

core where temperatures are much higher. This is an important direct empirical confirmation of neutrino trapping in dense matter, of great significance in determining supernova dynamics. One way to lower the mean ν_e energy and maintain good agreement with the data is to allow a non-thermal spectrum.

The time structure of the neutrino signal is consistent with several very different models. The data can be fit by a burst with time constant ≥ 1 s. This timescale is somewhat longer than some theoretical estimates, but is quite consistent with the order of magnitude expected from diffusive cooling processes at the edge of the neutrinosphere, and not with the emission of electron neutrinos associated with rapid deleptonization in the core. If all the observed events are supernova-related, then the models tested suggest that the later events are associated with a lower emission temperature. This is suggestive of cooling at late times. Better fits are obtained if some or all of the last three Kamiokanda events and perhaps the last IMB event are due to noise, but then there is poorer agreement with the expected isotropy of the signal.

Finally, three important areas of neutrino physics gain by this analysis of supernova data.

Neutrino mass. This analysis suggests that a conservative limit on the electron neutrino mass would be somewhat weaker than one might initially estimate. The consistency of a relatively long time constant for the neutrino burst, on the order of 5 s, suggests that a timescale of this order should be used to limit the dispersion in the neutrino signal due to finite mass effects. Moreover, our analysis suggests that the mean energies quoted by the

Kamiokanda group may be low, so that values at the upper end of the allowed range should probably be taken. Finally, there is reasonable evidence that the lowest-energy event(s) are due to noise. With these caveats in mind, even trivial dimensional-analysis arguments yield upper limits of $O(15 \text{ eV})$, as opposed to the value of $O(8 \text{ eV})$ which one might derive without them. Presumably any more detailed mass limit must take these results into account.

Neutrino species. Assuming that the neutrino luminosity is equally spread between different species, one can derive a bound on the number of light neutrinos. Theory suggests that the total neutrino luminosity is constrained to be $\leq 3 \times 10^{53}$ erg, the gravitational binding energy of a 1.4-solar-mass neutron star. The data (assuming no background events, three neutrino families, and a 50% enhancement of electron neutrinos over antineutrinos) is consistent with luminosities as small as $\sim 1.3 \times 10^{53}$ erg. This would rise to a value greater than the theoretical upper limit if there were more than seven light neutrino families. (This bound decreases to ~ 5 if the temperatures and luminosities that appear to fit the data best are used.) If two or three of the events are due to noise, then the lower luminosity bound decreases to $\sim 1 \times 10^{53}$ erg, and as many as nine light neutrino species (or equivalents) would be allowed. These are consistent with, but weaker than, bounds obtainable from terrestrial experiments and from cosmology.

Neutrino mixing. Finally, if we assume that muon and tau neutrinos have a higher emission temperature than electron neutrinos, then we may test for possible neutrino oscillations in vacuum on their voyage to Earth. There is in fact possible support for this in the Kamiokanda and IMB data. Good agreement is obtained for 3–4-MeV electron neutrinos and 7–8-MeV ‘muon’ neutrinos mixing at 30%, or more if muon and tau neutrino luminosities are slightly smaller. Assuming mass oscillations over the 50 kpc travel distance (and thus $m_\nu > 10^{-10} \text{ eV}$), such large mixings could resolve the solar neutrino problem^{9,10}.

As these results suggest, this first observation from the nascent field of extra-solar neutrino spectroscopy has provided an important new window on the Universe. Existing detectors have obtained a clear signal from a supernova. Theoretical models can now face empirical tests of consistency, which the standard predictions of Type II supernova neutrino emission seem to pass with ease. But although the results provide new insights into both supernova and neutrino physics, this analysis underscores the fact that the existing data do not allow detailed categorical conclusions to be drawn. On the other hand, even this first meagre observation indicates the potential power, both to particle physics and astrophysics, of using new particle detectors to investigate the heavens.

I thank Tran Thanh Van and the organisers of the Rencontres de Moriond, where I obtained pre-publication data directly from members of the four neutrino collaborations, and which stimulated this work. I thank M. Goldhaber for details of the IMB experiment, M. Turner for pointing out some omissions in the first version of this work (see also ref. 11), and J. Cooperstein, A. Guth, M. Davis and F. Wilczek for useful discussions. L.M.K. is a Visiting Scientist at the Harvard-Smithsonian Center for Astrophysics and a Presidential Young Investigator. This work was supported in part by DOE.

Received 5 June; accepted 10 September 1987.

1. Bethe, H., Yahil, A. & Brown, G. E. *Astrophys. J.* **262**, L7–L10 (1982).
2. Burrows, A. *Astrophys. J.* **283**, 848–852 (1984).
3. Wilson, J. R. *et al. Ann. N.Y. Acad. Sci.* **470**, 267–293 (1986).
4. Saavedra *et al. in The Standard Model, the Supernova 1987A* (ed. Tran Thanh Van, J.) (Editions Frontières, Gif-sur-Yvette, 1987).
5. Alexeyev *et al. in The Standard Model, the Supernova 1987A* (ed. Tran Thanh Van, J.) (Editions Frontières, Gif-sur-Yvette, 1987).

6. Hirata, K. *et al. Phys. Rev. Lett.* **58**, 1490–1493 (1987).
7. Bionta, R. M. *et al. Phys. Rev. Lett.* **58**, 1494–1496 (1987).
8. Cherry, M. L., Lande, K. & Fowler, W. (eds) *Solar Neutrinos and Neutrino Astronomy* (American Institute of Physics, New York, 1985).
9. Glashow, S. & Krauss, L. M. *Phys. Lett.* **B190**, 199–207 (1987).
10. Barger, V., Whisnant, K. & Phillips, R. J. N. *Phys. Rev.* **D24**, 538–541 (1981).
11. Kolb, E. W., Stebbins, A. & Turner, M. S. *Phys. Rev.* **D35**, 3598 (1987).
12. Bahcall, J. N. *Nature* **237**, 682–685 (1987).
13. Schramm, G. N. preprint, Univ. Chicago (1987).
14. Burrows, A. & Lattimer, J. M. preprint no. 87-10, Univ. Arizona theor. Phys. (1987).
15. Bludman, F. A. & Schinder, P. J. preprint no. UPR-0336T, Univ. Pennsylvania (1987).

A new case of gravitational lensing

J. Surdej*, P. Magain†, J.-P. Swings*, U. Borgeest‡, T. J.-L. Courvoisier§, R. Kayser‡, K. I. Kellermann||, H. Kühr¶ & S. Refsdal‡

* Institut d'Astrophysique, Université de Liège, Avenue de Cointe 5, B-4200 Cointe-Ougrée, Belgium

† European Southern Observatory, Casilla 19001 Santiago 19, Chile

‡ Hamburger Sternwarte, Gojenbergsweg 112, D-2050 Hamburg 80, FRG

§ ST-ECF, European Southern Observatory, Karl-Schwarzschild Strasse 2, D-8046 Garching bei München, FRG

|| National Radio Astronomy Observatory, Edgemont Road, Charlottesville, Virginia 22903-2475, USA

¶ Max-Planck-Institut für Astronomie, Königstuhl, D-6900 Heidelberg 1, FRG

Even before the discovery of the first gravitational lens system¹ in 1979 (see ref. 2 for a recent review of the other published candidates), it was recognized that a statistical evaluation of the occurrence of gravitational lensing within a well-defined sample of quasars is important to understand better the quasar luminosity function and possibly the QSO phenomenon itself³, to test cosmological models^{4,5} and to probe the luminous and dark matter distribution on various scales in the Universe⁴⁻⁶. A year ago, we began a systematic search from the European Southern Observatory (ESO) at La Silla, Chile, for gravitational lens systems in a selected sample of highly luminous quasars; $M_V < -29.0$. These objects are promising candidates for gravitationally lensed QSO images with arc-second and sub-arc-second separations (J.S. *et al.*, manuscript in preparation and refs 7 and 8). Since December 1986, we have identified four possible cases. Here we give a brief description of our first identified gravitational lens system UM673 = Q0142-100 = PHL3703⁹⁻¹¹. It consists of two images, A ($m_R = 16.9$) and B ($m_R = 19.1$), separated by 2.2 arc s at a redshift $z_q = 2.719$. The lensing galaxy ($m_R \approx 19$, $z_L \approx 0.49$) has also been found. It lies very near the line connecting the two QSO images, ~ 0.8 arc s from the fainter one. Application of gravitational optometry to this system leads to a value $M_0 \approx 2.4 \times 10^{11} M_\odot$ for the mass of the lensing galaxy and to $\Delta t \approx 7$ weeks for the most likely travel-time difference between the two light paths to the QSO (assuming $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0$).

Our optical search for gravitationally lensed highly luminous quasars (HLQs) is being performed with a charge-coupled device (CCD) camera at the Cassegrain focus of the ESO/MPI (Max-Planck Institut) 2.2-m telescope. The selected HLQs are first observed through narrow-band filters chosen to isolate bright redshifted emission lines (Ly α or occasionally C III) as well as nearby continua in order to search for lensed QSO image candidates at moderate angular separations (≥ 1 arc s). Taking advantage of the large dynamic range of the CCD detectors and of the excellent resolution of the 2.2-m telescope under optimal seeing conditions, we are able to resolve secondary images up to 4 magnitudes fainter than the primary one at angular separations near 1.1–1.3 arc s. We show in Fig. 1a the central part of a direct CCD frame obtained for UM673 on 6 December 1986 with the ESO faint object spectrograph and camera (EFOSC)¹² at the $f/8$ Cassegrain focus of the ESO 3.6-m telescope. The CCD image was taken during a 2-min exposure through a Bessel R filter, and it clearly shows that UM673 is spatially resolved into two star-like images A ($m_R = 16.9$) and B ($m_R = 19.1$), separated by 2.2 arc s. The diffuse object labelled D in Fig. 1a is a galaxy in the field of the CCD.

Low dispersion spectra of UM673 A and B covering several spectral ranges from 3,660 to 10,000 Å were also taken with EFOSC at the beginning of December 1986. The B300 spectra recorded in the 3,800–7,040 Å wavelength interval at a resolution (FWHM) of ~ 13 Å are reproduced in Fig. 2. Further description

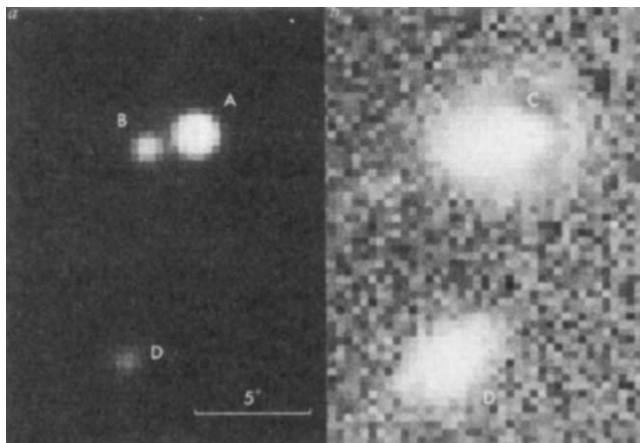


Fig. 1 a, The central part of a CCD frame of UM673 obtained with EFOSC in a 2 min exposure through a Bessel R filter on 6 December 1986. The highly luminous quasar is resolved into two quasi-stellar images A ($m_R = 16.9$) and B ($m_R = 19.1$) separated by 2.2 arc s. b, The result of subtracting the images of UM673 A and B from the original frame (a) after modelling their shapes with a double gaussian profile. The residual image is that of the lensing galaxy ($m_R \approx 19$) at a redshift $z_L \approx 0.49$. Object D may be a member of an associated cluster of galaxies. On both these figures, the size of a single pixel is 0.338". North is up and East to the left. Courtesy of ESO.

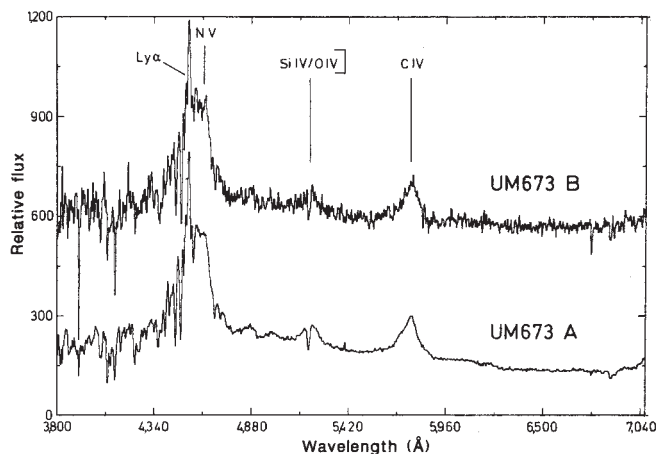


Fig. 2 The B300 low-dispersion spectra of UM673 A and B observed in December 1986 with EFOSC at the $f/8$ Cassegrain focus of the ESO 3.6-m telescope. The spectrum of UM673 B has been multiplied by a factor 8.8 and offset by 400 units in ordinate. In this and the next figure, the ordinates refer to a relative flux scale and the abscissae correspond to wavelengths in angstroms.

of the instrumental setting and reduction and analysis of these and additional spectroscopic data will be published elsewhere. A careful comparison of the redshifts ($z_q(A) = z_q(B) = 2.719 \pm 0.001$), derived from the Ly α + NV, Si IV/O IV] and C IV emission lines, as well as the perfect superposition achieved between the broad emission line (see the peculiar Ly α + NV blend) and narrow absorption line profiles in the spectra of UM673 A and B clearly establish an exceptional physical similarity between the two QSO images. It is therefore suggested that UM673 A and B are the gravitationally lensed images of a single distant quasar. This assertion is further supported by the result of a cross-correlation between the B300 spectra which leads to a velocity difference $\Delta v = -24 \pm 109 \text{ km s}^{-1}$ between the QSO images A and B.

Furthermore, the spectrum of UM673 B presents a small but clear excess of continuum radiation at wavelengths ≥ 5800 Å.

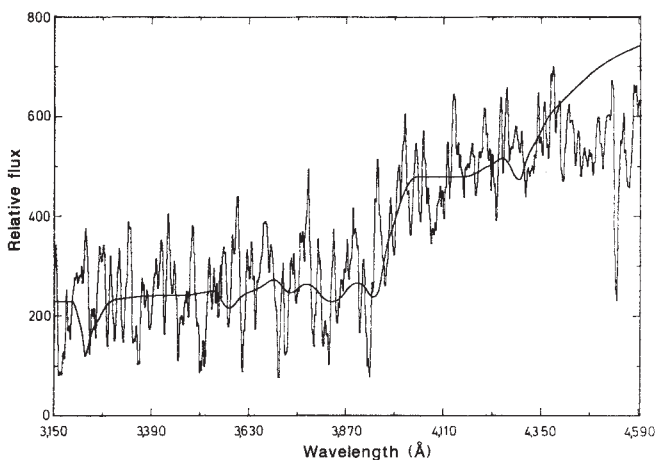


Fig. 3 The mean spectrum of brightest galaxies observed in distant galaxy clusters¹³ is represented in this figure by the smooth line. It is a weighted superposition of 48 reasonably well-exposed spectra that have been cleaned and shifted to zero velocity. Note the prominent 4,000 Å break in this mean spectrum. The noisy spectrum is the subtraction of the B300(A) from the B300(B) spectrum after a flux normalization in the spectral range 3,800–5,700 Å and correction to zero redshift, assuming that this residual spectrum is that of the lensing galaxy at $z_L = 0.49$.

Assuming that the apparent reddening of UM673 B is caused by light contamination from an additional object, we extracted this excess of radiation as follows. We first multiplied the B300(B) spectrum by a constant factor so that the B300(B) and B300(A) spectra perfectly match each other in the wavelength range 3,800–5,700 Å. We then subtracted these two normalized spectra to obtain a residual spectrum that is well matched by the mean spectrum of brightest galaxies in distant clusters¹³ at a redshift $z = 0.49 \pm 0.01$ (see Fig. 3). We identify this residual spectrum as that of the gravitational lens producing the two observed QSO images.

A careful processing of the R CCD frame shown in Fig. 1a allowed us to detect the image of the suspected lensing galaxy. We show in Fig. 1b the result of subtracting the two QSO images from the original CCD frame (Fig. 1a) after modelling their shapes with a double gaussian profile. The heights of these profiles have been adequately scaled according to the QSO image brightnesses. We find that the centroid of the lensing galaxy is located very close to the line joining the two QSO images, 0.8 arc s from the position of UM673 B. The integrated magnitude of this galaxy is measured to be approximately $m_R = 19$, a value comparable to what is predicted for an elliptical galaxy at a redshift $z_g = 0.5$ (ref. 14).

Because we do not have yet any definite information on the exact orientation of the lensing galaxy, or on its immediate surroundings (whether there is an associated galaxy cluster, for example), we have used a simple model to estimate the mass M_0 of the lensing galaxy (within the projected images) and the expected delay Δt between the travel times of the two lensed QSO images. Our calculations are based on a lens model made of a single galaxy that is axially symmetric with respect to the line of sight¹⁵ and use the observed redshifts and positions of UM673 A and B, and of the deflector. For deceleration parameter $q_0 = 0$, $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and adopting a smooth mass distribution in the lensing galaxy, we find $M_0 \approx 2.4 \times 10^{11} M_\odot$ and $\Delta t \approx 7$ weeks. From an observational point of view, this predicted time delay appears ideal as it should be possible to measure it accurately within a single observing season. The challenge is great because it should allow, in principle, a determination of the galaxy mass M_0 to be made independently of the value of Hubble's constant H_0 ¹⁶. UM673 also provides an excellent case

of a gravitational lens for which it should be possible to model the lensing mass distribution in detail. This may then allow an independent determination of certain cosmological parameters.

We thank Marc Azzopardi, Valentina Zitelli and Hugo Schwarz for making available some of their observing time. J.S. is a Chercheur Qualifié au Fonds National de la Recherche Scientifique (Belgium), and T.J.-L.C. is affiliated to the Astrophysics Division of the Space Science Department of ESA.

Received 20 July; accepted 12 August 1987.

1. Walsh, D., Carswell, R. F. & Weymann, R. J. *Nature* **279**, 381–384 (1979).
2. Canizares, C. R. in *IAU Symp. 124: Observational Cosmology* (eds Hewitt, A., Burbidge, G. & Fang, L. Z.) 729–744 (Reidel, Amsterdam, 1987).
3. Barnothy, J. M. & Barnothy, M. F. *Science* **162**, 348–352 (1968).
4. Refsdal, S. *Mon. Not. R. astr. Soc.* **128**, 307–310 (1964).
5. Refsdal, S. *Mon. Not. R. astr. Soc.* **132**, 101–111 (1966).
6. Canizares, C. R. *Nature* **291**, 620–624 (1981); erratum *Nature* **293**, 490 (1981).
7. Turner, E. L., Ostriker, J. P. & Gott, J. R. III *Astrophys. J.* **284**, 1–22 (1984).
8. Ostriker, J. P. & Vietri, M. *Astrophys. J.* **300**, 68–76 (1986).
9. MacAlpine, G. M. & Feldman, F. R. *Astrophys. J.* **261**, 412–421 (1982).
10. Berger, J. & Fringant, A.-M. *Astr. Astrophys. Suppl.* **61**, 191–202 (1985).
11. Véron-Cetty, M.-P. & Véron, P. *European Southern Observatory Scient. Rep.* no. 5 (1987).
12. Dekker, H. & D'Odorico, S. *European Southern Observatory Operating Manual* no. 4 (1985).
13. West, R. M. & Kruszewski, A. *Irish astr. J.* **15**, 25–35 (1981).
14. Coleman, G. D., Wu, C.-C. & Weedman, D. W. *Astrophys. J. Suppl. Ser.* **43**, 393–416 (1980).
15. Borgeest, U. & Refsdal, S. *Astr. Astrophys.* **141**, 318–322 (1984).
16. Borgeest, U. *Astrophys. J.* **309**, 467–471 (1986).

Sustained magnetic fields in binary millisecond pulsars

G. Chanmugam & K. Brecher*

Space Telescope Science Institute, Homewood Campus, Baltimore, Maryland 21218, USA and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA
* Department of Astronomy, Boston University, Boston, Massachusetts 02215, USA

The two known binary millisecond pulsars^{1,2} PSR1953+29 and PSR1855+09 with periods $P = 6.1 \text{ ms}$ and 5.4 ms , respectively, and the long orbital-period binary radio pulsars³ are believed to have evolved from low-mass X-ray binaries (LMXBs)⁴. The neutron stars in these binaries are believed to have formed from the accretion-induced collapse of white dwarfs with shorter orbital periods⁵. These neutron stars have weak magnetic fields ($3 \times 10^8 \text{ G} \leq B \leq 6 \times 10^{11} \text{ G}$) and rapid spin ($5 \text{ ms} \leq P \leq 1 \text{ s}$), compared with typical isolated pulsars, and are believed to have been spun up by accretion after their magnetic fields had decayed (see ref. 6 for review) on a timescale $t_D \sim 9 \times 10^6 \text{ yr}$ (ref. 7). If correct, the fraction of pulsars which are binary millisecond pulsars is too large⁸ by a factor of >100 . This severe discrepancy is removed if, as we propose here, the magnetic fields of neutron stars do not decay either in such binaries or in general. We also show that, if such neutron stars are formed from the accretion-induced magnetic flux and angular momentum-conserving collapse of white dwarfs, most of them are likely to have been born, and remain, spinning rapidly and to have weak magnetic fields, in agreement with observations of binary millisecond pulsars and LMXBs.

It is widely believed that the neutron stars in the seven known binary radio pulsars (see Table 2 of ref. 3) were spun up during an earlier evolutionary phase when the neutron star was accreting matter from its companion star and was an X-ray pulsar. If there is adequate time to reach equilibrium an accreting neutron star spins at the equilibrium period for disk accretion⁹:

$$P_{\text{eq}} = 6 \text{ ms } B_9^{6/7} R_6^{18/7} (M/M_\odot)^{-5/7} \dot{M}_{17}^{-3/7} \quad (1)$$

where B_9 is the neutron star's magnetic field in units of 10^9 G , R_6 its radius in units of 10^6 cm , M its mass and $\dot{M}_{17}$ the accretion rate in units of 10^{17} g s^{-1} .

It is also believed that neutron stars in binaries are born with magnetic fields $B \sim 10^{12}$ G and spin relatively slowly ($P \gg 1$ ms), as is also believed to be the case for isolated pulsars. The field is then assumed to decay¹⁰ on a timescale t_D . This leads to a decrease in P_{eq} (spin-up), and when accretion stops a fast radio pulsar is formed.

If the field continues to decay then within $\sim 8 t_D$ its value becomes $< 5 \times 10^8$ G, characteristic of the millisecond pulsars, and soon the pulsar becomes unobservable. The short lifetime of the millisecond pulsars and the paucity (~ 50) of their progenitor LMXBs imply⁸ that the number of binary ms pulsars observable should be $\ll 1$, contrary to observations. Worse discrepancies arise for the other binary radio pulsars which have higher ($\sim 10^9$ – 10^{11} G) fields and shorter lifetimes. To overcome this difficulty it has been suggested that the magnetic field stops decaying^{8,11}, or t_D increases, when the field gets down to $\sim 10^9$ – 10^{11} G. On the other hand, many of the galactic bulge sources, which show no X-ray pulsations, probably have even weaker ($\leq 10^8$ G) fields, indicating that field decay continues to occur down to a lower field. Thus it is entirely unclear why for different neutron stars, in binaries, field decay stops at widely different levels, making such assumptions *ad hoc*.

Here we therefore make the much simpler hypothesis that the magnetic fields of neutron stars formed from accretion-induced collapse of white dwarfs do not decay at all. This is, of course, satisfied if neutron star magnetic fields do not decay in general. But then why do these neutron stars have relatively weak fields, compared to isolated pulsars? To examine this and the related question of field decay, let us first discuss the origin of magnetic fields of isolated neutron stars. There appear to be two views: (1) the field is produced by flux conservation when the core of an evolved massive star collapses to produce a neutron star^{12,13}; or (2) the field is created by thermoelectric effects in the crust of the neutron star¹⁴ after formation. Assuming the latter mechanism there are difficulties in explaining the existence of magnetic fields in young pulsars¹⁵, such as the Crab pulsar, and the creation of dipolar fields. It has been claimed that an attractive feature of this model is that it predicts field decay on a timescale of $\sim 10^6$ yr as a result of ohmic dissipation of the crustal magnetic field. Yet, the first detailed numerical calculations of this decay show that the field does not decay exponentially as is assumed in analyses of pulsar observations¹⁶. This is because, in addition to diffusing outwards, the field diffuses inwards to regions of higher electrical conductivity. Furthermore, for initial fields of 10^{12} G, the field does not decay down to millisecond pulsar fields even on the Hubble time. It is therefore very likely that (1) holds, in which case, if field decay occurs, it must take place through instabilities^{17–20}. In this case too the timescale for decay is given by that for the dissipation of crustal currents or magnetic fields. Thus there is no satisfactory explanation of field decay if (1) does occur. Although it is widely believed that magnetic fields of neutron stars decay, alternative suggestions have been made^{21,22} and there is no consensus^{6,23}. The lack of any reasonable mechanism for field decay supports the view that magnetic fields of neutron stars do not decay.

If the magnetic fields of neutron stars are produced from flux conservation then the fields of neutron stars in LMXBs will depend on the magnetic fields of the progenitor white dwarfs which, in general, may be different from the fields in the cores of massive stars. Nothing is known observationally about the fields of such inner cores but there is now a significant amount of observational data on white dwarfs. Among the cataclysmic variables (CVs) which contain accreting white dwarfs, about 20% have significant magnetic fields²⁴. These are divided into two classes, the AM Her²⁵ and DQ Her²⁶ binaries, each presently containing about 10 systems²⁷. In the AM Her binaries the magnetic field of the white dwarf $B_{WD} \sim \text{few} \times 10^7$ G and the white dwarf rotates synchronously. In the DQ Her binaries the white dwarf rotates asynchronously as may be deduced from

the optical and X-ray pulsations they exhibit. For the magnetic field to channel the accreting matter onto the magnetic poles it is necessary that the magnetospheric radius r_m exceeds the radius of the star so that typically one requires $B \gtrsim 10^4$ G. In no case has the field been determined spectroscopically. In only one case (3A0729+103) has any optical polarization been detected²⁸ implying $B \sim 10^7$ G if it is due to cyclotron emission²⁹. Thus the fields in magnetic CVs known so far probably lie in the range $10^4 \leq B \leq 10^7$ G. Note that there is no evidence of field decay among white dwarfs which are isolated or in binaries and that the AM Her binaries, which have strong fields, have short orbital periods, indicating that they are old²⁷. Thus the suggestion⁸ that accretion leads to field decay in neutron stars does not work when applied to white dwarfs.

The accretion-induced collapse of a white dwarf³⁰ in a CV, assuming flux conservation, results in a neutron star with a field $B \approx B_{WD}(R_{WD}/R)^2$. Hence if $R_{WD} \sim 5 \times 10^8$ cm and $R \sim 10^6$ cm, in roughly 80% of the cases $B_{WD} \leq 10^4$ G so that $B \leq 2.5 \times 10^9$ G. The LMXBs which show no pulsations and hence probably have weak fields and the binary millisecond pulsars probably belong to this class. In $\sim 20\%$ of the cases neutron stars with $B \sim 2.5 \times 10^9$ G to 7.5×10^{12} G would be produced, consistent with the fields in the binary radio pulsars which are descendants of the LMXBs. These are crude estimates as not all CVs produce neutron stars. Whether the mass of a white dwarf increases or not due to mass transfer depends on the rate of accretion. Low rates of accretion lead to nova explosions which result in no increase in the mass of the white dwarf⁴. Thus few CVs are likely to produce neutron stars. Fortunately, the number of active LMXBs in our galaxy is ≤ 100 whereas there are $\sim 2 \times 10^5$ CVs in our galaxy^{4,31}, so that only a small fraction of the CVs need produce neutron stars.

If magnetic field decay does occur among isolated pulsars it is likely that this is due to instabilities. In this case it may depend on the structure and shape of the magnetic field^{19,20}. Neutron stars which are descendants of white dwarfs with stable magnetic fields may have the appropriately stable field structures already built into them and hence will remain stable if these properties are unaffected by the supernova explosion leading to the formation of the neutron star. This may not be the case for neutron stars which are produced from the cores of massive stars.

The rotation periods of white dwarfs in the DQ Her binaries^{26,27} range from $P_{rot} \sim 30$ s to $\sim 10^4$ s with the majority having $P_{rot} \leq 10^3$ s. The synchronously rotating white dwarfs in the AM Her binaries have $5 \times 10^3 \leq P_{rot} \leq 1.3 \times 10^4$ s by virtue of their orbital motion. White dwarfs with weaker ($B \leq 10^4$ G) fields are also likely to be spinning rapidly ($P_{rot} \leq 10^3$ s) because either they were born spinning rapidly or they have been spun-up by boundary-layer accretion³². The accretion-induced collapse of a white dwarf is likely to produce a neutron star spinning with a period $P = P_{WD}(I/I_{WD})$, where $I \sim 10^{45}$ g cm² and $I_{WD} \sim 10^{50}$ g cm² are the moments of inertia of the neutron star and white dwarf, if angular momentum is conserved. Hence one would expect $3 \times 10^{-4} \text{ s} \leq P \leq 0.1$ s so that the neutron star would be born spinning rapidly. There is no difficulty in producing millisecond pulsars, in fact one needs in extreme cases to get rid of angular momentum³³ to produce a neutron star spinning more slowly than its break-up speed. Note that by comparison the spin of a neutron star formed from the collapse of the core of an evolved isolated massive star would reflect its possibly slower rotation. On the other hand a similar core which collapses in a binary is likely to be spinning more rapidly because of interaction with the common envelope formed in an earlier phase³⁴. Such a collapse is likely to form a more rapidly spinning neutron star such as, for example in the short orbital-period binary radio pulsar PSR1913+16 which contains a 60-ms pulsar.

It is likely that the weak-field neutron stars would typically be born spinning more rapidly than the strong field ones, reflecting the equilibrium period for disk accretion (equation (1)) for the progenitor white dwarfs. Once born the weak-field neutron

stars will continue to spin rapidly, as is the case for the binary millisecond pulsars. If the magnetic field of the progenitor white dwarf, and hence neutron star, is strong then after formation the neutron star will spin down, during the accretion phase, to a relatively long equilibrium period, if there is no field decay.

An example of a strong-field neutron star is Her X-1, for which $B \sim 5 \times 10^{12}$ G (ref. 35). It is believed to be an old binary (age $\sim 10^9$ yr) but may contain a relatively young neutron star formed from the accretion-induced collapse of a white dwarf. It is assumed that the binary was ejected from near the galactic plane, as a result of the asymmetric momentum released during the collapse, and that it reached its present position several kpc above the galactic plane in $\sim 10^7$ yr (ref. 5). The mass ejection during the formation of the neutron star would have led to a sudden increase in the orbital separation so that the companion star no longer filled its Roche lobe. After $\sim 2-3 \times 10^7$ yr the companion would have filled its Roche lobe again, the system becoming an X-ray source. If field decay takes place, the field of the neutron star should during this time have decayed by a factor $\geq e^3$, contrary to the strong field detected, unless its initial field was much stronger than that of a typical pulsar. These difficulties do not exist if the magnetic field does not decay.

The mechanism discussed here cannot easily explain the formation of the isolated 1.6-ms pulsar PSR1973+16 unless the companion was disrupted³⁶ or merged⁵ with the white dwarf which produced the neutron star. Thus it remains possible that it was formed by itself³⁷. On the other hand, a binary origin for the probably isolated 3-ms pulsar recently discovered³⁸ in the globular cluster M28 is more likely as the binary could have been disrupted by stellar collisions in the dense stellar environment of the globular cluster³⁹.

We have emphasized that neutron stars formed from the accretion-induced collapse of white dwarfs must have properties characteristic of their white dwarf progenitors which are, in general, different from the cores of evolved massive stars which are the progenitors of isolated neutron stars. Many of these should hence be born with relatively weak magnetic fields and be spinning rapidly in good agreement with observations of the LMXBs and binary millisecond pulsars if their fields do not decay.

G.C. thanks Icko Iben for useful discussions. This research was supported by an NSF grant and carried out when G.C. was on sabbatical leave at Space Telescope Science Institute.

Received 12 June; accepted 9 September 1987.

- Boriakoff, V., Bucccheri, R. & Fauci, F. *Nature* **304**, 417-419 (1983).
- Segelstein, D. J., Rawley, L. A., Stinebring, D. R., Fruchter, A. S. & Taylor, J. H. *Nature* **322**, 714-717 (1986).
- Dewey, R. J., Maguire, C. M., Rawley, L. A., Stokes, G. H. & Taylor, J. H. *Nature* **322**, 712-714 (1986).
- McClintock, J. & Rappaport, S. A. in *Cataclysmic Variables and Low-Mass X-Ray Binaries* (eds Lamb, D. Q. & Patterson, J.) 61-77 (Reidel, Dordrecht, 1985).
- Taam, R. E. & van den Heuvel, E. P. J. *Astrophys. J.* **305**, 235-245 (1986).
- Taylor, J. H. & Stinebring, D. R. *Rev. Astr. Astrophys.* **24**, 285-327 (1986).
- Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613-639 (1985).
- van den Heuvel, E. P. H., van Paradijs, J. A. & Taam, R. E. *Nature* **322**, 153-155 (1986).
- Savonije, G. J. *Nature* **325**, 416-418 (1987).
- Gunn, J. & Ostriker, J. P. *Astrophys. J.* **160**, 979-1002 (1970).
- Kulkarni, S. R. *Astrophys. J.* **306**, L85-L89 (1986).
- Ruderman, M. A. & Sutherland, P. G. *Nature phys. Sci.* **246**, 93 (1973).
- Levy, E. M. & Rose, W. K. *Nature* **250**, 40-41 (1974).
- Blandford, R., Applegate, J. & Hernquist, L. *Mon. Not. R. astr. Soc.* **204**, 1025-1048 (1983).
- Radhakrishnan, V. in *Birth and Evolution of Neutron Stars: Issues Raised by Millisecond Pulsars* (eds Reynolds, S. P. & Stinebring, D. R.) 337-341 (NRAO, Charlottesville, 1984).
- Sang, Y. & Channugam, G. *Astrophys. J.* (in the press).
- Flowers, E. & Ruderman, M. *Astrophys. J.* **215**, 302-310 (1977).
- Wang, Z. & Eichler, D. *Astrophys. J.* (submitted).
- Taylor, R. J. *Mon. Not. R. astr. Soc.* **162**, 17-23 (1971).
- Channugam, G. in *Birth and Evolution of Neutron Stars: Issues Raised by Millisecond Pulsars* (eds Reynolds, S. P. & Stinebring, D. R.) 213-217 (NRAO, Charlottesville, 1984).
- Kundt, W. *Astr. Astrophys.* **98**, 207-210 (1981).
- Beskin, V. S., Gurevich, A. V. & Istomin, Ya. N. *Astrophys. Space Sci.* **102**, 310-326 (1983).
- Ruderman, M. A. in *Highlights of Modern Astrophysics* (eds Shapiro, S. L. & Teukolsky, S. A.) 21-44 (Wiley, New York, 1986).
- Morris, S. L. *et al. Astrophys. J.* **314**, 641-652 (1987).
- Liebert, J. & Stockman, H. S. in *Cataclysmic Variables and Low-Mass X-Ray Binaries* (eds Lamb, D. Q. & Patterson, J.) 151-177 (Reidel, Dordrecht, 1985).
- Lamb, D. Q. & Patterson, J. in *Cataclysmic Variables and Related Objects, Proc. IAU Colloq. No. 72* (eds Livio, M. & Shaviv, G.) 229-237 (Reidel, Dordrecht, 1983).
- Channugam, G. & Ray, A. *Astrophys. J.* **285**, 252-257 (1984).
- Penning, W., Schmidt, G. D. & Stockman, H. S. *Astrophys. J.* **301**, 881-885 (1986).

- Channugam, G. & Frank, J. *Astrophys. J.* (in the press).
- Canal, R. & Schatzman, E. *Astr. Astrophys.* **46**, 229-235 (1976).
- Warner, B. *Mon. Not. R. astr. Soc.* **167**, 61P-64P (1974).
- de Kool, M. & van Paradijs, J. *Astr. Astrophys.* **173**, 279-283 (1987).
- Pacini, F. *Astr. Astrophys.* **126**, L11-L12 (1983).
- Iben, I. & Tutukov, A. V. *Astrophys. J. Suppl. Ser.* **58**, 661-710 (1985).
- Trumper, J. *et al. Astrophys. J.* **219**, L105-L110 (1978).
- Ruderman, M. & Shahan, J. *Comm. Astrophys.* **10**, 15-22 (1983).
- Brecher, K. & Channugam, G. *Nature* **302**, 124-125 (1983).
- Middleditch, J. *et al. IAU Circ. No.* 4401 (1987).
- Krolik, J. H., Meiksin, A. & Joss, P. C. in *Cataclysmic Variables and Low-Mass X-Ray Binaries* (eds Lamb, D. Q. & Patterson, J.) 107-116 (Reidel, Dordrecht, 1985).

Occultation observations of compact radio sources through comet Halley's plasma tail

S. Ananthakrishnan, P. K. Manoharan & V. R. Venugopal

Radio Astronomy Centre, Tata Institute of Fundamental Research, PO Box 8, Udhagamandalam 643 001, India

Enhancement in scintillations of radio sources attributable to plasma irregularities in the cometary tail have been reported for comet Kohoutek¹ and recently for comet Halley^{2,3}. The radio sources suitable for studying the plasma tail of a comet must have a significant fraction of their flux density in the sub-arc second structure, so a set of 30 sources lying within 2° of the path of comet Halley and with a flux density of $S_{408} \geq 0.8$ Jy were chosen from the Molonglo Catalogue⁴ and interplanetary scintillation (IPS) observations of them were made with the Ooty Radio Telescope at 327 MHz during January 1986⁵. From these, four were found to be suitable for cometary scintillation observations; these were done between February and April 1986. We monitored nearly simultaneously other scintillating sources outside the tail for comparison and, in contrast with earlier claims of positive detection, found no significant increase in the level of turbulence that could be attributed to the plasma tail.

The four radio sources observed had scintillating structure of size ≤ 0.15 arc s and scintillating flux density ΔS in the range of $\sim 0.3-5$ Jy. They were: 2052-106 ($\Delta S \approx 0.5$ Jy), 2021-168 ($\Delta S \approx 0.3$ Jy), 1921-293 ($\Delta S \approx 5$ Jy) and 1817-391 ($\Delta S \approx 1.2$ Jy). The last source is confused with 1815-391 in the Ooty telescope beam and hence IPS may arise from either or both of them. In addition other IPS sources near each of the above sources but lying outside the comet's tail during occultation were selected as control sources for monitoring purposes.

Each of the occulted sources along with its nearby control sources was observed over 3 days centred on the day of occultation. The observations were spread over 6-9 h during which the occulting source was observed in stretches of 15-20 min alternated between 10 min of observation of each of the control sources. A nearby cold region of the sky was also observed to establish an 'off source' reference approximately once every

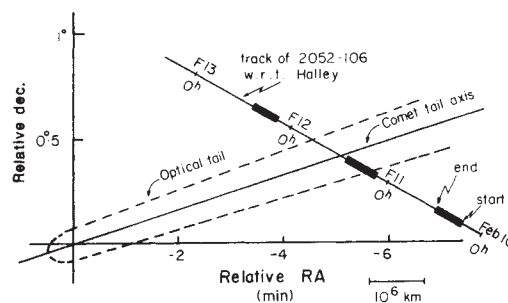


Fig. 1 Track of the radio source 2052-106 with respect to the nucleus of the comet Halley. The coordinates given are relative right ascension and declination.

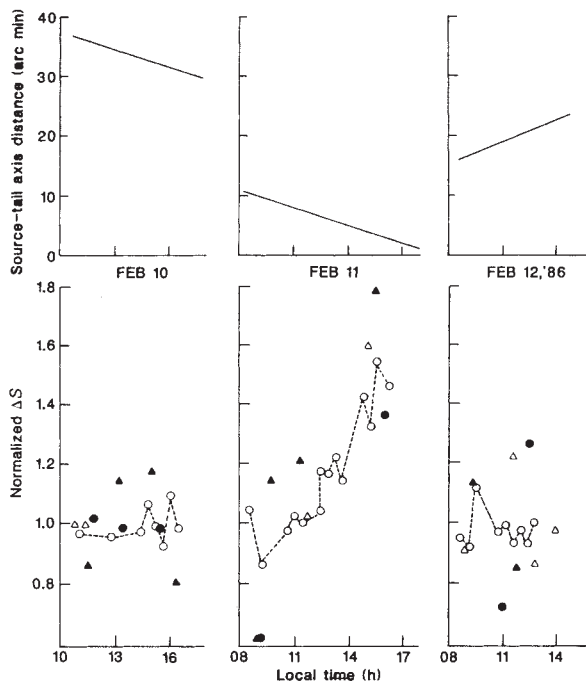


Fig. 2 The scintillating flux density of 2052–106 normalized to its average value for the day is plotted as a dashed line in the lower panel. The same measure for the comparison sources 2052–071, 2053–201 and 2052–097 are marked $\blacktriangle$, $\triangle$ and $\bullet$, respectively. The distance from the central axis of the plasma tail to the line of sight to the source 2052–106 is shown in the upper panel.

hour. The receiver bandwidth was 4 MHz and the time constant for data acquisition on magnetic tapes was 0.05 s. The data were sampled every 20 ms. A time constant of 0.5 s was used for recording on chart, mainly for interference monitoring. The digital data were analysed following normal procedures for IPS observations⁶. The power spectra were computed by standard fast Fourier transform algorithms and corrected for the 'off source' noise. The r.m.s. intensities were computed from the integrals under the corrected power spectra.

Figure 1 shows the geometry of observation for 2052–106 on 11 February 1986. This takes into account the small offset of $\sim 1.5^\circ$ caused by the topocentric speed of the comet of $\sim 12.5 \text{ km s}^{-1}$. Three more scintillating sources with ΔS between 100 and 200 mJy were observed as control sources. These were 2052–071, 2053–201 and 2052–097. Figure 2 shows the normalized ΔS for 2052–106 and the control sources for 10, 11 and 12 February 1986. We have also plotted the perpendicular distance of the source from the line of sight to 2052–106 to the axis of the tail of the comet. It is seen that ΔS for 2052–106 increased significantly on the day of occultation (11 February) as the plasma tail approached the source. One might have concluded that the observed rise in scintillations on 11 February was due to the comet, had there been no observations of the control sources. But the three control sources, similarly normalized, show an equal increase around the same time. On 10 and 12 February there were no significant increases on any of the sources. These data demonstrate the importance of having control sources for such cometary occultation observations.

1817–391 was observed on 1 April 1986. The results for this source and its control source 1827–360 are shown in the same format in Fig. 3a. During the first half of the observing period of $\sim 9 \text{ h}$, there was no detectable scintillation from 1817–391, while 1827–360 being a much stronger source had detectable scintillating flux; but during the second half both 1817–391 and 1827–360 showed an increase of ΔS .

The observations of 2021–168 and 1921–293 on 2 and 24 March, respectively, showed no significant variations

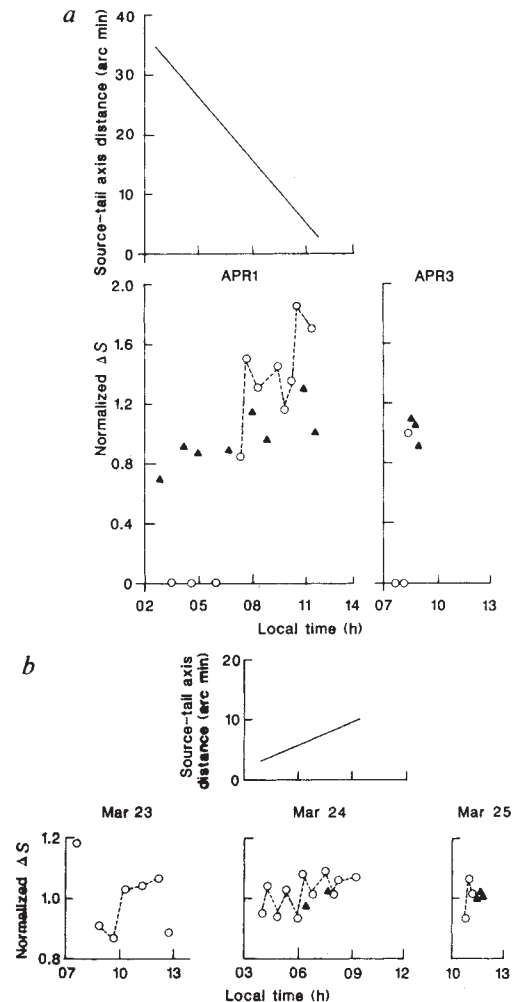


Fig. 3 Similar format to that of Fig. 2. In *a*, 1817–391 and its control 1827–360 are marked by $\circ$ and $\blacktriangle$, respectively. In *b*, 1921–293 and its control 1938–15 are marked by $\circ$ and $\blacktriangle$, respectively.

throughout the periods of occultation. The signal-to-noise ratio on 2021–168 was not high enough to set a reasonable bound but 1921–293 was useful. The latter source and its control, 1938–15, are shown in Fig. 3b.

Thus while observing four scintillating radio sources through the tail of comet Halley during its 1986 apparition, on only two occasions was an apparent increase noticed in the scintillations that might have been attributed to the comet, but at both these, similar increases were observed in the scintillation on the nearby control sources. From a careful analysis of the power spectral data on these two days, 11 February and 1 April 1986, it appears that whereas the former could be due to an increase in the solar-wind turbulence, the latter is more probably due to ionospheric scintillations. Day-to-day variations in normalized scintillating flux density due to the solar wind are $\sim \pm 20\%$ as seen in Figs 2 and 3 but these variations may be up to a factor of three.

Turbulence in space plasmas, such as the solar wind, the ionosphere and the interstellar medium is typically characterized by power-law spectra. Although the action of the Fresnel filter confines scintillation phenomena to scale sizes $\leq 400 \text{ km}$ it is known that the changes in IPS are the effect of the conditions within the interplanetary plasma as a whole including the presence of corotating structures. Hence, we have considered the ratio of the level of the electron-density spectrum $\Phi_n(K)$ to the mean-squared electron density $\langle n_e^2 \rangle$ as a measure of the plasma turbulence rather than inferring $(\Delta n_e)^2$ from IPS observations. In the quiet solar wind, Φ_n/n_e^2 is constant over a large elongation

range. Because the scintillating flux density variance, $\Delta S^2 \propto \Phi_{ne} L$, where L is the effective thickness of the intervening medium, the increase of the normalized variance, $\Delta S^2 / \langle \Delta S^2 \rangle$, can be written

$$\frac{\Delta S^2}{\langle \Delta S^2 \rangle} = \frac{((\Phi_{ne}/n_e^2)_s n_{es}^2 L_s + (\Phi_{ne}/n_e^2)_c n_{ec}^2 L_c)}{(\Phi_{ne}/n_e^2)_s n_{es}^2 L_s}$$

where the subscripts 's' and 'c' refer to the solar wind and cometary plasma respectively and $\langle \Delta S^2 \rangle$ is the estimated variance due to the solar wind on that day. From our observations we find that the limit on $\Delta S^2 / \langle \Delta S^2 \rangle$ is < 0.4 . Estimates of this ratio and of n_{ec}^2 , L_c , n_{es}^2 , L_s are given in Table 1. From the last column in Table 1 it is seen that the value of Φ_{ne}/n_e^2 in the plasma tail did not exceed its mean value in the solar wind for our observations. If L_c were assumed to be 0.5×10^6 km, all the values in the last column would increase by a factor of two.

In contrast to our results two recent observations of enhanced scintillations caused by comet Halley have been reported by Alurkar *et al.*² and Slee *et al.*³ at 103 and 408 MHz respectively. Both groups have reported enhancements in scintillation index by a factor of 4–5 (equal to the square root of the values in column 1 of Table 1). They converted these into estimates of Δn_e^2 using gaussian models for the turbulence spectrum. We have used the data provided by both groups to estimate Φ_{ne}/n_e^2 by comparing the observed ΔS with what is normally expected at such elongations. These estimates are shown in Table 1.

The latter two observations of scintillation enhancement shown in Table 1 represent remarkable increases in the observed degree of turbulence. But in the absence of control source observations, it is desirable to draw any conclusions judiciously.

The increase in scintillations reported by Alurkar *et al.*² has been confirmed by the Japanese IPS group for 327 MHz. But this group had been observing a grid of sources regularly and they attribute the observed increase during the occultation by comet Halley to an unfortunately located corotating interaction region (M. Kojima, personal communication).

Slee *et al.*³ have reported that their occultation source 1827–360 had scintillation index, $m_{408} \approx 0.016$ on 27, 28 and 31 March and have compared the increase in scintillation observed on 29

the plasma tail of a comet (P. Duffett-Smith, personal communication⁹). These results confirm the conclusions in our paper.

We thank Dr A. Pramesh Rao, Professor W. A. Coles, Professor M. N. Joshi and Professor G. Swarup and other colleagues at the Radio Astronomy Centre for their valuable suggestions and criticisms. We also thank Dr M. Kojima, Nagoya University, Japan for giving permission to quote his results before publication. Our grateful thanks to Dr Peter Duffett-Smith for his valuable comments and suggestions.

Received 6 May; accepted 10 July 1987.

1. Ananthakrishnan, S., Bhandari, S. M. & Pramesh Rao, A. *Astrophys. Space Sci.* **37**, 275–283 (1975).
2. Alurkar, S. K., Bhonsle, R. V. & Sharma, A. K. *Nature* **322**, 493–441 (1986).
3. Slee, O. B., McConnel, D., Lim, J. & Bobra, A. D. *Nature* **325**, 699–701 (1987).
4. Large, M. I. *et al. Mon. Not. R. astr. Soc.* **194**, 693–704 (1981).
5. Swarup, G. *et al. Nature* **230**, 185–188 (1971).
6. Rao, A., Pramesh, Bhandari, S. M. & Ananthakrishnan, S. *Aust. J. Phys.* **27**, 105–120 (1974).
7. Reme, H. *et al. Nature* **321**, 349–352 (1986).
8. Grard, R. *et al. Nature* **321**, 290–291 (1986).
9. Hajivassiliou, C. A. & Duffett-Smith, P. J. *Mon. Not. R. astr. Soc.* (in the press).

Possible explanation for the superconducting 240-K phase in the Y–Ba–Cu–O system

H. Büttner & A. Blumen

Physikalisches Institut, University of Bayreuth,
D-8580 Bayreuth, FRG

A recent report¹ on the reverse alternating current Josephson effect in the $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ system strongly indicates a superconducting phase in this material with a transition temperature at ~ 240 K. Another report has been published on possible superconductivity at 230 K in the Eu-compound². The structure of these phases and how they might be related to the 90-K phase is not known. Here we show that high-temperature superconductivity may be envisaged as a fractal structure whose localized vibrations are responsible both for an effective electron–electron coupling and a high transition temperature T_c . This model should also be applicable to the La–Ba–Cu–O system where the transition temperature of the disordered phase shows superconductivity at ~ 35 K while structural instabilities are observed up to 200 K^{3,6}.

Most of the recent theoretical explanations of high-temperature superconductivity start from regular structures and use the conventional BCS-mechanism^{7,8}, resonating valence bond⁹, plasmon-type coupling (ref. 10; J. Ashkenazi, C. G. Kuper and R. Tyk, preprint), bipolaron mechanism¹¹, negative U-effects¹² or Peierls instabilities¹³ to explain the phenomenon. Our considerations are prompted by the experimental observation reported both in the first high- T_c paper³ and also in many subsequent investigations, that the superconducting state may be of percolative nature^{14,15}, possibly existing only on phase boundaries^{16–18}. We therefore suggest that the high-temperature superconducting state is supported by a partial (or total) fractal lattice and that the effective electron–electron coupling is mediated by the localized vibrational states (fractons). Because then the electrons at the Fermi edge are still in an extended (band) state (see ref. 18), we assume that the electron–fracton coupling is similar to the electron–phonon coupling; therefore only the density of states $F(\omega)$ is relevant for the effective coupling. For the transition temperature, however, a second parameter, namely the relevant frequency ω_0 of the phonon modes is very important, as can be seen from the approximate expression for T_c first discussed by McMillan^{19,20}

$$T_c \approx \omega_0 \exp \left(-\frac{1+\lambda}{\lambda - \mu^*} \right) \quad (1)$$

Table 1 Parameters for the plasma tail of comet Halley

Reference	ΔS^2 $\langle \Delta S^2 \rangle$	n_{es}^2 (cm ³) (ref. 7)	n_{ec}^2 (cm ³) (ref. 8)	L_s (10 ⁸ km)* (10 ⁵ km)	L_c (10 ⁵ km)	$(\Phi_{ne}/n_e^2)_c$ $(\Phi_{ne}/n_e^2)_s$
This work	< 0.4	5	30	0.9	1.0	< 1
Alurkar <i>et al.</i> ²	27	5	10	1.5	1.0	975
Slee <i>et al.</i> ³	16	5	30	1.6	1.0	75

* Estimated assuming spherical symmetry.

March ($m_{408} = 0.062$) with this value. But the same source was observed fortuitously by us on 1–3 April as a control source (Fig. 3) and we find that its $m_{327} = 0.038$, which is about twice as high as the first value given above, taking into account the frequency difference in observations. Taking our observation as the reference, the estimated value of the turbulence variance for the observations by Slee *et al.*³ would be reduced by a factor of five. This would make their observed increase similar to that of Figs 2 and 3. Because they made no observations of control sources it is difficult to say whether their observed increase should properly be attributed to the comet's plasma tail.

We conclude that no increase in the level of turbulence clearly attributed to the plasma tail of the comet has been demonstrated by these IPS observations. The sensitivity of our observation was such that, had there been any marked increase in the normalized level of turbulence Φ_{ne}/n_e^2 in the comet tail over that in the solar wind we would certainly have detected it.

Since submission of this manuscript we have learnt of a recent study of ~ 40 occultations by ~ 20 comets that failed to reveal a significant increase in the level of turbulence attributable to

where λ describes the coupling to the phonons and can be calculated from

$$\lambda = 2 \int d\omega \alpha^2 F(\omega) / \omega \quad (2)$$

with $F(\omega)$ being the density of states, while α^2 is the coupling function²¹. The parameter μ^* describes the Coulomb repulsion and ω_0 is proportional to the Debye frequency ω_D of the material.

We compare a conventional translation-invariant structure with a percolation fractal structure. The first class of materials is characterized by a density of states $F_{ph}(\omega) \approx \omega^2$, a coupling constant α^2 , a high-frequency cutoff $\omega_0 \approx \omega_D$ and a Coulomb parameter μ^* . The fractal structure, on the other hand, has a density of states $F_f(\omega) \approx \omega^{d-1} \approx \omega^{1/3}$ where d is the fracton dimension²² (and for percolative clusters $d = 4/3$). For parameter λ the low frequencies are important and, assuming for the moment the same ω_D (although the highest frequencies are different, see below) and the same ω -independent α^2 , one has

$$\begin{aligned} F_{ph} &= 3\omega^2 / \omega_D^3 \rightarrow \lambda_{ph} = 3\alpha^2 / \omega_D \\ F_f &= 4\omega^{1/3} / 3\omega_D^{4/3} \rightarrow \lambda_f = 8\alpha^2 / \omega_D \end{aligned} \quad (3)$$

which is a drastic increase in coupling strength for the fractal structure. Furthermore, the relevant point for an increase in the transition temperature is that in a fractal structure (with the same number of sites as in the conventional material) the highest possible frequencies are not the same as assumed above but are much higher than in the nonfractal structure^{23,24}. In an example on a Sierpinski-structure Rammal²⁴ had shown that

$$\omega_0 = 2(d+1)\omega_D \quad (4)$$

where d is the euclidean dimension. For a two-dimensional Sierpinski-gasket this will give a factor of six compared to the conventional structure. This factor increases further in three-dimensional networks, or if one includes more than one displacement per site²⁵ T_c will increase by a comparable factor if the coupling to the high-frequency phonons is as effective as indicated above. Realistic models would also show phonon-fracton crossovers with a comparable increase in the highest frequencies. For a detailed comparison with experiments such effects must be included but we do not do this here.

These effects which lead to an increase in ω_0 and λ_f may explain the high-temperature phases in the La-Ba-Cu-O materials and its yttrium-derivative. The reported structural instabilities may also result from this percolative structure which will be very sensitive to chemical and physical processing.

We note that additional granular effects due to Josephson contacts may further enhance T_c ²⁶.

It would be interesting to investigate the mechanism suggested here by experimental studies of materials with varying degrees of disorder (which may change d and also ω_0). Moreover, we note that different experiments monitor different aspects, such that the conductivity is a property of the bulk backbone, whereas the Meissner effect is a probe of the surface layer.

The work was supported by the Deutsche Forschungsgemeinschaft (SFB 213) and by the Fonds der Chemischen Industrie.

Received 8 June; accepted 22 September 1987.

- Chen, J. T., Wenger, L. E., McEwan, C. J. & Logothetis, E. M. *Phys. Rev. Lett.* **58**, 1972-1975 (1987).
- Huang, C. X. *et al. Nature* **328**, 403-404 (1987).
- Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189-193 (1986).
- Geiser, U. *et al. Phys. Rev. B* **35**, 6721-6724 (1987).
- Paul, D. McK. *et al. Phys. Rev. Lett.* **58**, 1976-1978 (1987).
- Fossheim, K. *et al. Solid St. Commun.* **63**, 531-533 (1987).
- Matheiss, L. F. *Phys. Rev. Lett.* **58**, 1028-1031 (1987).
- Weber, W. *Phys. Rev. Lett.* **58**, 1371-1374, 2154 (E) (1987).
- Anderson, P. W. *Science* **235**, 1196-1198 (1987).
- Ruvalds, J. *Phys. Rev. B* **35**, 8869-8872 (1987).
- Prelovsek, P., Rice, T. M. & Zhang, F. C. *J. Phys. C* **20**, L229-L233 (1987).
- Zhang, L. Y. *Solid St. Commun.* **62**, 491-493 (1987).
- Machida, K. & Kato M. *Phys. Rev. B* **36**, 854-856 (1987).
- Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908-910 (1987).
- Sulewski, P. E. *et al. Phys. Rev. B* **35**, 5330-5333 (1987).

- Ginley, D. S. *et al. Phys. Rev. B* **36**, 829-832 (1987).
- Sun, J. Z. *et al. Phys. Rev. Lett.* **58**, 1574-1576 (1987).
- Entin-Wohlman, O., Alexander, S. & Orbach, R. *Phys. Rev. B* **32**, 8007-8012 (1985).
- McMillan, W. L. *Phys. Rev.* **163**, 331-341 (1968).
- Allen, P. B. in *Dynamical Properties of Solids* Vol. 3 (eds Horten, G. K. & Maradudin, A. A.) 95-196 (North-Holland, Amsterdam, 1980).
- Marsiglio, F. & Carbotte, J. P. *Solid St. Commun.* **63**, 419-423 (1987).
- Aharony, A., Alexander S., Entin-Wohlmann, O. & Orbach, R. *Phys. Rev. B* **31**, 2565-2567 (1985).
- Alexander, S. *et al. Phys. Rev. B* **28**, 4615-4619 (1983).
- Rammal, R. *J. Phys., Paris* **45**, 191-206 (1984).
- Yu, K. W. *Physica* **128A**, 307-317 (1984).
- Chakravarty, S., Kivelson, S., Zimanyi, G. T. & Halperin, B. I. *Phys. Rev. B* **35**, 7256-7259 (1987).

Synthesis of the high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ in small particle size

A. Manthiram* & J. B. Goodenough

Center for Materials Science and Engineering, ETC 5.160,
The University of Texas at Austin, Austin, Texas 78712-1084 USA

Technological exploitation of the recently discovered high- T_c superconducting oxides requires processing the materials at modest temperatures. A first step toward this goal is the synthesis of the material in small particle size. This requires, in turn, the design of a synthetic route at temperatures below those where sintering takes place. Here we report the preparation of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ as small particles by firing the coprecipitated oxalates at temperatures $T \leq 780^\circ\text{C}$; sintering begins above 800°C . Properties of the materials depend on the highest temperatures reached. As noted independently by Torardi *et al.*¹, low-temperature synthesis in air results in a tetragonal, semiconducting product rather than the orthorhombic, superconducting phase. Therefore we also report a low-temperature ($T < 780^\circ\text{C}$) method of converting the tetragonal phase to the superconducting phase without sintering of the small particles

Coprecipitation of yttrium, barium and copper as oxalates in stoichiometric proportions was achieved by the following procedure. BaCO_3 (with 14 wt% excess) and Y_2O_3 were dissolved in hot, dilute nitric acid. Cupric nitrate (with 17 wt% excess) was then dissolved in the solution before cooling to room temperature. The excesses of BaCO_3 and $\text{Cu}(\text{NO}_3)_2 \cdot 2\frac{1}{2}\text{H}_2\text{O}$ were necessary to obtain stoichiometric proportions in the precipitate. An 80 wt% excess of 1 M oxalic acid was added to the cooled solution and the pH was adjusted to 1.5 by adding dilute ammonia. The pale-green slurry formed was allowed to stand for ~ 1 h, filtered on a sintered-glass funnel, washed with distilled water and dried at 100°C . The resulting powder was well ground before firing to decompose the oxalates and form the desired oxide.

The oxalate powders were fired in air and furnace-cooled; the products formed were examined by X-ray powder diffraction. Single-phase products were annealed in an oxygen atmosphere at 450°C for 15 h and then 350°C for 12 h followed by slow cooling in the furnace. This annealing procedure has been found optimal for intercalating the maximum oxygen content into the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ structure²⁻⁴. The oxygen content was determined by iodometric titration as described elsewhere². The oxygen contents and lattice parameters of several oxygen-annealed phases obtained at different initial firing temperatures, are given in Table 1.

The maximum initial firing temperature required for complete reaction of the oxalates was found to be $\sim 780^\circ\text{C}$; five days at this temperature with one intermittent grinding gave nearly complete reaction with traces ($< 2\%$) of unreacted BaCO_3 , Y_2O_3 and CuO . The oxygen content of this phase was found to correspond to $\text{YBa}_2\text{Cu}_3\text{O}_{6.70}$, lower than the maximum content

* Permanent address: Madurai Kamaraj University, Madurai, India.

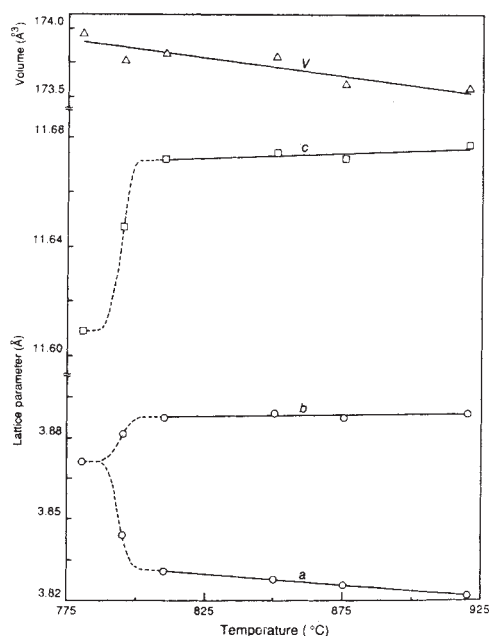


Fig. 1 Variation of lattice parameter with initial firing temperature for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ formed by oxalate decomposition followed by an oxygen anneal.

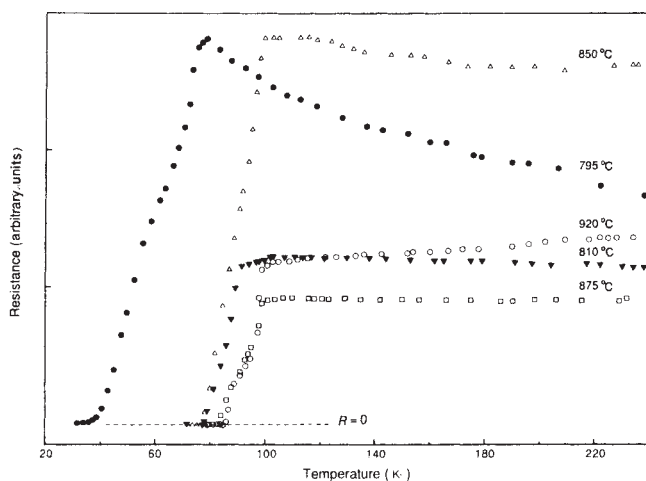


Fig. 2 Temperature variation of resistance for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ pellets formed from oxalate decomposition after an oxygen anneal for different oxalate-decomposition temperatures in $^{\circ}\text{C}$.

6.96 obtained under the same oxygen anneal for the orthorhombic, superconducting phase having a sharp superconducting transition temperature $T_c \approx 90 \text{ K}^{2-4}$. Moreover, the phase is semiconducting (not superconducting) and its X-ray powder diffraction could be indexed on a tetragonal unit cell (Table 1) with $c \approx 3a$. The $\text{YBa}_2\text{Cu}_3\text{O}_{6.70}$ composition having an orthorhombic symmetry is superconducting with $T_c \approx 55 \text{ K}^{4,5}$.

The tetragonal phase has the same basic structure as the orthorhombic, superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ phase, but the oxygen in the Cu(1) planes (the copper planes lying between Ba planes) are disordered on the a - and b -axes rather than being preferentially ordered onto b -axis sites. The tetragonal lattice parameters given in Table 1 are in close agreement with those found independently by Torardi *et al.*¹ by neutron diffraction for an impure phase of high oxygen content prepared at low temperature.

The formation of a tetragonal phase with such high oxygen concentration is in sharp contrast to the orthorhombic, superconducting phase obtained by intercalating oxygen into a tetragonal phase of low (6.0–6.2) oxygen content obtained at

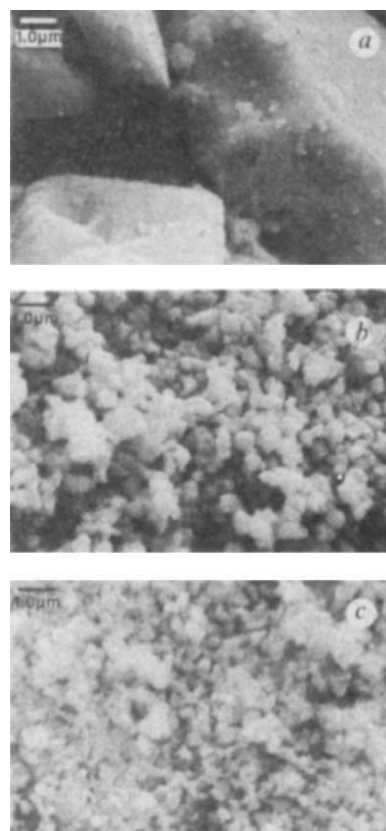


Fig. 3 Scanning electron microscope photographs for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ prepared *a*, by usual ceramic method at 920°C ; *b*, at 810°C from oxalate precursor; and *c*, by heating tetragonal $\text{YBa}_2\text{Cu}_3\text{O}_{6.70}$ in N_2 at 750°C , all followed by O_2 annealing at 450 – 350°C .

900 – $950^{\circ}\text{C}^{2-4}$. The oxygen of the Cu(1) layer intercalate reversibly under these conditions, and ordering of the oxygen onto the b -axis sites occurs for oxygen contents > 6.3 at room temperature. Oxygen is lost progressively with increasing temperature, and an orthorhombic $\rightleftharpoons$ tetragonal transition occurs at $T_i \approx 650^{\circ}\text{C}$ where the oxygen content is ~ 6.5 ^{3,6}. The observation of a tetragonal phase with oxygen content ≥ 6.70 at room temperature, reluctant to intercalate oxygen to higher concentrations, indicates that the oxygen of the Cu(1) layer is not only disordered, but also of a different character.

One of us⁷ has argued that the presence of disordered oxygen beyond a critical oxidation state ($\delta < 0.5$), which can occur during synthesis at lower temperatures, results in formation of the less mobile peroxide ion $(\text{O}_2)^{2-}$ in the Cu(1) layer. Consistent with this hypothesis is the observation (Table 1 and Fig. 1) that oxalates fired above 800°C form single phases that, after oxygen annealing, have the superconducting orthorhombic structure with the higher oxygen contents needed for $T_c \approx 90 \text{ K}$. From the progressive contraction of the a parameter in Fig. 1, the degree of ordering is seen to increase with the initial firing temperature.

The resistances of all samples were measured on pressed pellets, subjected to the oxygen anneals, by a standard four-probe technique. Whereas the tetragonal sample was semiconducting down to 4.2 K , all the other samples were superconducting (Fig. 2) with T_c increasing with the initial firing temperatures (Table 1).

Particle sizes were determined with a JEOL JSM-35C scanning electron microscope (SEM). Figure 3*a* and *b* compare the SEM photographs taken for the superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ oxide prepared at 920°C by standard ceramic techniques and at 810°C from an oxalate precursor. The oxalate method gives a soft, fine

Table 1 Oxygen contents and lattice parameters of several oxygen-annealed phases at different initial firing temperatures

T (°C)	t (h)	$(7-\delta)$	Lattice parameter a	b	c	V (Å ³)	T_c (K)
780	120	6.70	3.871 (2)		11.609 (5)	173.96	—
795	96	6.75	3.844 (2)	3.881 (2)	11.647 (3)	173.76	60*
810	96	6.87	3.831 (1)	3.887 (1)	11.672 (3)	173.81	87
850	72	6.89	3.828 (1)	3.889 (1)	11.674 (2)	173.79	89
875	48	6.91	3.826 (1)	3.887 (1)	11.672 (3)	173.58	92
920	36	6.93	3.822 (1)	3.889 (1)	11.677 (2)	173.56	95

Initial firing schedule (temperature T and time t) for oxalate decomposition to form $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ and after subsequent oxygen anneal, product oxygen content $(7-\delta)$, lattice parameters, unit-cell volume V and superconducting transition temperature T_c

* Broad transition with onset at 78 K and $\rho = 0$ at 40 K.

powder and shows smaller particle sizes. Nevertheless, it is apparent that sintering has been initiated in the samples prepared at 810 °C. To avoid sintering, it is necessary to devise a strategy for converting the tetragonal phase, prepared at 780 °C, to the orthorhombic structure.

As the orthorhombic phase is obtained from a tetragonal phase of lower oxygen content, it was decided to anneal the 780 °C product in an inert atmosphere at <780 °C before subjecting it to an oxygen anneal. The tetragonal sample was heated in N_2 at 750 °C for 16 h and quenched; it had a reduced oxygen

content of 6.20 and an X-ray powder-diffraction pattern similar to that for tetragonal $\text{YBa}_2\text{Cu}_3\text{O}_6$. Subjecting this sample to the O_2 anneal resulted in an oxygen content of 6.86 and an orthorhombic X-ray pattern. Moreover, the particle size of this superconducting ($T_c = 90$ K) sample remained small, see Fig. 3c, with little evidence of the initiation of sintering.

We thank Dr A. L. de Lozanne and Mr S. H. Pan for providing facilities and assistance for the resistance measurements and Dr M. A. Schmerling for recording the SEM photographs. This research was supported by the R. A. Welch Foundation, Houston, Texas.

Received 8 September; accepted 23 September 1987.

1. Torardi, C. C. *et al.* *Chemistry of High-Temperature Superconductors*, ACS Symposium Series 351 (eds Nelson, D. L., Whittingham, M. S. & George, T. F.) 152-163 (Am. Chem. Soc., Washington DC, 1987).
2. Manthiram, A., Swinnea, J. S., Sui, Z. T., Steinfink, H. & Goodenough, J. B. *J. Am. chem. Soc.* (in the press).
3. Gallagher, P. K., O'Bryan, H. M., Sunshine, S. A. & Murphy, D. W. *Mater. Res. Bull.* **22**, 995-1006 (1987).
4. Tarascon, J. M. *et al.* *Proc. Int. Workshop on Novel Mechanisms of Superconductivity Berkeley*, 22-26 June, 1987 (in the press).
5. Batlogg, B. *et al.* *Proc. Int. Workshop on Novel Mechanisms of Superconductivity Berkeley*, 22-26 June, 1987 (in the press).
6. Schüller, I. K. *et al.* *Proc. Int. Workshop on Novel Mechanisms of Superconductivity Berkeley*, 22-26 June, 1987 (Plenum, New York, in the press).
7. Goodenough, J. B. (in preparation).

Atmospheric iridium at the South Pole as a measure of the meteoritic component

Gurdal Tuncel & William H. Zoller*

Middle East Technical University, Department of Environmental Engineering, Ankara, Turkey

The measurement of iridium (Ir) in atmospheric samples is important because it gives information on the short-term flux of extraterrestrial material without interference from fragmentation products from large bodies. Such information cannot be obtained from sediment samples, because sediment samples integrate over millions of years and include contributions of large bodies impacted in that time period. In addition to flux information, through the analysis of Ir in atmospheric samples we can also evaluate a possible contribution of extraterrestrial material to the unusual enrichment of chalcophilic elements in a remote atmosphere. We present here a determination of average particle-borne Ir concentration in the South Pole atmosphere. The average values of $(7.3 \pm 3.1) \times 10^{-17} \text{ g m}^{-3}$ suggests that the concentration of extraterrestrial material in the South Pole atmosphere is not large enough to explain the enrichments of anomalously enriched elements; however, meteoritic material contributes significantly to the observed concentrations of Co, Fe and Mn. We estimate an accretion rate for background extraterrestrial material of 11,000 tons annually.

We analysed Ir in 30 atmospheric particulate samples collected at the South Pole during 1979. Samples were collected on 25-cm-diameter cellulose (Whatman) filters at the Amudsen Scott station at the geographical South Pole. Sampling was continuous, being interrupted only to change filters. During each 10-day sampling period, approximately 20,000–30,000 m^3 of air was passed through the filter. Half of each Whatman filter was used for Ir analysis, and the remaining halves for analyses for other trace elements by instrumental neutron activation analysis (INAA) and sulphate by ion-chromatography.

Samples were first counted for natural radioactivity of ^{70}Be , then irradiated at the US National Bureau of Standards (NBS)

Table 1 Mass concentrations of components in the South Pole atmosphere in 1979

Component	Summer (ng m^{-3})	Winter (ng m^{-3})
Sulphate	192 ± 125	97 ± 39
Marine	26 ± 13	95 ± 52
Crustal	7.4 ± 4.1	3.9 ± 4.2
Enriched	1.1 ± 0.5	0.84 ± 0.31
Meteoritic	0.18 ± 0.14	0.14 ± 0.08

Uncertainties are standard deviations per sample observation. Data from ref. 5, except for the meteoritic component which is calculated from measured Ir concentration by assuming chondritic composition.

reactor at a neutron flux of $5 \times 10^{13} \text{ cm}^{-2} \text{ s}^{-1}$ for 8 hours. After 1–2 months of decay, samples were opened, dissolved in hot fuming HNO_3 , and a known amount of Ir carrier was added. After repeated heating with small portions of H_2O_2 , Ir was removed from the cellulose matrix by coprecipitation with $\text{Fe}(\text{OH})_3$. The precipitate was dissolved in concentrated hydrochloric acid and base-metals were removed by cation exchange. The Ir in the eluent from the cation-exchange step was collected on anion-exchange resin, which was air-dried and sealed in a polyethylene bag. Spectra of γ -rays emitted by samples were collected for about 2 days with a $\text{Ge}(\text{Li})$ or intrinsic Ge γ -ray detector. The chemical yield of the separation procedure is $70 \pm 10\%$. To determine the yield, a known amount of Ir stock solution was pipetted on to cellulose filters, filters were irradiated, counted, separated and counted for a second time. We were able to measure Ir without radiochemical separation in several samples collected between 1980 and 1983. The yield determined by counting two of these samples before and after separation were 64% and 72%, which showed that the dissolution of Ir and losses during the separation procedure were same for both samples and standards.

The Ir concentration in the South Pole atmosphere varies from 2×10^{-17} to $13.4 \times 10^{-17} \text{ g m}^{-3}$. It is higher during austral summer than in winter, with average concentrations of 9.1×10^{-17} and $7.2 \times 10^{-17} \text{ g m}^{-3}$, respectively. The Ir mass measured in eight blank filters, which were similar cellulose filters placed on the sampling system for few minutes were, 0.49, 0.24, 0.14, 0.26, 0.43, 0.33, 0.28 and 0.26 pg per filter with an average of

* Present address: Department of Chemistry, University of Washington, Seattle, Washington 98195, USA.

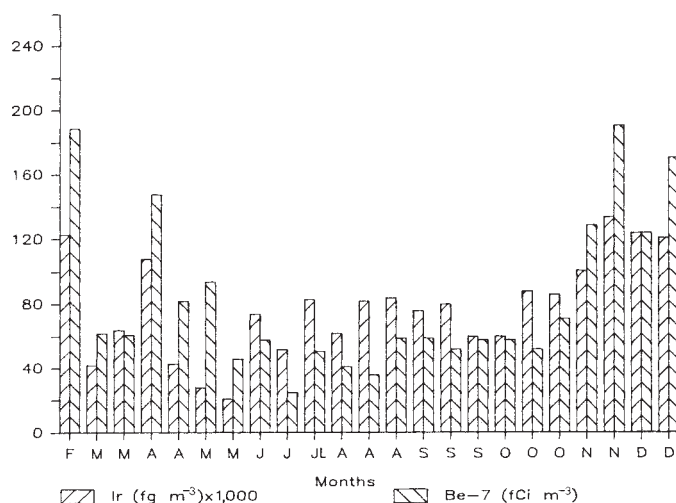


Fig. 1 Variation of ^7Be activity and Ir concentration in the South Pole atmosphere ($\text{fg} = 1 \times 10^{-15} \text{ g}$, $\text{fCi} = 1 \times 10^{-15} \text{ Ci}$). Number of samples taken in each month are not necessarily the same.

$0.3 \pm 0.1 \text{ pg}$. The average Ir mass in blank filters was approximately 30% of the average Ir mass measured in the sample filters.

Figure 1 shows Ir concentrations and ^7Be activities in the South Pole samples. Beryllium-7 is mostly produced in the stratosphere by the interaction of cosmic rays with atmospheric nitrogen¹. Earlier it was shown that variations in the ^7Be activity in the South Pole atmosphere are closely related to similar fluctuations in the activities of fission products produced by the nuclear detonations in the stratosphere^{2,3}. Therefore, any relation between Ir and ^7Be may indicate a stratospheric source for Ir.

Iridium concentrations and ^7Be activity in the South Pole atmosphere are strongly correlated ($r=0.75$), suggesting a stratospheric source for Ir. The most likely source of Ir in the stratosphere is extraterrestrial material. Although large Ir enrichments are observed in emissions from the Kilauea volcano in Hawaii⁴, temporal variations of Ir in the South Pole atmosphere did not show any correlation with elements such as As, Sb and Zn, which are believed to have a volcanic origin in the South Pole atmosphere^{2,5}. We believe that Ir concentrations can be used to estimate both the meteoritic component in the background aerosols and the global accretion rate of extraterrestrial material.

Known components of atmospheric aerosol at the South Pole are listed in Table 1. Concentrations of components are determined by applying chemical mass balances (CMBs) to 30 samples collected during 1979⁵. The meteoritic component accounts for very little of the aerosol mass at the South Pole atmosphere. The CMBs indicate that extraterrestrial materials can account for 12% of Co, 6% of Fe and 4% of Mn in the atmosphere, and less than 1% of other elements. Extraterrestrial material obviously cannot explain unusual enrichments of chalcophilic elements such as As, Se, Zn, Cu, Ag, Cd, In, Sb and Au in the South Pole atmosphere.

Table 2 shows estimates of the meteorite accretion rate reported in the literature. Estimated flux values range from 10^2 to $10^7 \text{ tons yr}^{-1}$. Parkin and Tilles⁶, reviewing estimates reported before 1967, concluded that the global flux of extraterrestrial material converges to about $100,000 \text{ tons yr}^{-1}$. The large differences in reported accretion rates are partly due to inherent limitations of individual techniques used in estimating the flux, such as size selectivity, the use of markers that are also produced by terrestrial, natural and anthropogenic sources, and difficulties in associating the measured marker with the meteorite flux. Studies which involve core samples from various oceanic sediments also suffer from some common difficulties, such as uncertainties in

the noble metal content of crustal material⁷, possible contribution of Kilauea volcano to the Ir concentrations measured in the Pacific basin, and uncertainties in the sedimentation rate.

Estimates of extraterrestrial flux based on the polar ice and snow samples are less prone to such uncertainties. Earlier studies at the South Pole have shown that flux of crustal dust is one-to-two orders of magnitude smaller than the similar flux at mid-latitudes, which reduces the contribution by terrestrial Ir². Using Al concentrations measured in the South Pole atmosphere (370 pg m^{-3}), Taylor's crustal abundances⁸ and 0.025 parts per 10^9 as the crustal abundance of Ir⁹, we estimate that the concentration of terrestrial (crustal) Ir in the South Pole atmosphere is $1.4 \times 10^{-19} \text{ g m}^{-3}$, which is two orders-of-magnitude smaller than the observed Ir concentration.

To estimate the global flux of meteoritic material, we first converted atmospheric Ir concentrations to the corresponding concentrations in snow. The relationship between the atmosphere and snow can be found by comparing measured concentrations of trace elements in atmospheric particulate and snow samples. Snow-to-air ratios for crustal and marine elements are not the same. The ratio for elements with marine origin, such as Na, K, Mg and Ca are 0.4, 0.3, 0.3 and 0.4, respectively; however, the ratio for typical crustal elements, such as Al, Mn and Fe, are 2.1, 2.7 and 1.6, respectively, and that for S is $1.0^{5,10,11}$. Thus, crustal particles are scavenged more efficiently from the atmosphere than marine elements and sulfate at the South Pole, in agreement with Kumai's electron microscope studies on ice crystals in the South Pole atmosphere¹².

The meteoritic material collected on our filters consists of either small, unablated, unattached silicate grains or ablation products. The snow-to-air ratio for unablated silicate grains is expected to have a similar snow-to-air ratio to that of crustal dust (2.13). However, ablation of meteorites produces ultrafine particles, which are picked up by sulphate aerosol in the stratosphere or upper troposphere¹³. For this attached meteoritic component, the snow-to-air ratio should be similar to that for sulphate (1.0). In our calculations, we assumed that the snow-to-air ratio for extraterrestrial material in the South Pole atmosphere is between 1.0 and 2.13.

By using the range of 1.0–2.13 as the snow-to-air ratio for the Ir at the South Pole, the expected concentration of Ir in snow is calculated as 8.0×10^{-17} to $1.6 \times 10^{-16} \text{ g per g of snow}$. The density of surface snow at the South Pole is 0.4 g cm^{-3} (ref. 14) and the snow-accumulation rate is 17 cm yr^{-1} (ref. 15). Using these values, the local Ir flux is calculated as 5.4×10^{-16} to $1.1 \times 10^{-15} \text{ g cm}^{-2} \text{ yr}^{-1}$. Assuming a chondritic composition, in which the Ir concentration is 0.51 p.p.m.¹⁶, we obtain a meteorite flux of 1.1×10^{-9} to $2.1 \times 10^{-9} \text{ g cm}^{-2} \text{ yr}^{-1}$ and a global flux of 6×10^3 to $11 \times 10^3 \text{ tons yr}^{-1}$, the latter assuming the same deposition flux over the entire surface of the Earth.

As the source of extraterrestrial material is in the upper atmosphere, the latitudinal distribution of Ir in the lower atmosphere and Earth's surface depends on mixing between the stratosphere and troposphere, which occurs mostly at mid-latitudes and is minimal in polar regions. Due to this uneven distribution of extraterrestrial material on the Earth's surface, meteorite flux values based on polar regions must be multiplied by a highly uncertain constant to obtain an average over all latitudes. Brocas and Picot¹⁷ estimated the constant as 2.0 from the latitudinal distribution of ^{90}Sr , but later values vary between 2.0 and 4.0^{18,19} as in case of ^{90}Sr , ^7Be is also produced in the stratosphere and, consequently, should show the similar trends of mixing between the stratosphere and troposphere. Beryllium-7 activity measured at the South Pole is 80 fCi m^{-3} and its average activity at eight stations over the Pacific Ocean is 60 fCi m^{-3} (ref. 20). Beryllium-7 does not indicate any preferential deposition of stratospheric particles at midlatitudes. Due to the uncertainty in this constant, we did not multiply our flux value by any factor.

Cunningham and Zoller² estimated the meteorite flux as 600

Table 2 Reported estimates of global meteorite accretion rate

Method used in estimate	Accretion rate (tons yr ⁻¹)	Ref.
Magnetic spherules from Pacific clay.	90	25
²⁶ Al in Greenland ice	<100,000	26
²⁶ Al in Pacific sediments	1.3 × 10 ⁶	27
⁵³ Mn in Antarctic ice	100,000	18
Iron in particles collected from stratosphere	<90,000	28
Ni in Pacific sediments	(1–2) × 10 ⁶	29
	400,000	30
Ir in Mn nodules	20,000	31
	70,000	32
Ir in Pacific sediments	110,000	9
	78,000	23
Ir in Antarctic ice	400,000	19
Ir in Antarctic and Greenland ice	10,000–20,000	21
Ir in South Pole aerosol	6,000–11,000	This work
Co in Antarctic atmosphere	600	2
³ He/ ⁴ He ratio in Pacific sediments	2,000	33
Satellite, visual and radio-meteor studies	16,000	22

tons yr⁻¹ by using the Co concentration in the South Polar aerosol. This unusually small flux estimate probably results from assumptions Cunningham and Zoller used to convert atmospheric meteorite concentration to a global flux value. When our model is applied to Co concentrations reported by Cunningham and Zoller, a flux value of 26,000 tons yr⁻¹ is obtained. Temporal variations of Co at the South Pole suggest that Co may have a source other than crustal material and meteorites, probably volcanic or anthropogenic⁵. Consequently, the accretion rate estimated from Co concentrations is an upper limit rather than true flux value.

Takahashi *et al.*²¹ measured the Ir concentration in the Antarctic snow as 3.1 × 10⁻¹⁶ g per g snow and estimated the meteorite flux as (10–20) × 10³ tons yr⁻¹, both of which agree with our estimates. Ganapathy also measured the Ir concentration in the South Polar snow layers¹⁹, obtaining an estimated flux of 100,000 tons yr⁻¹. (He then multiplied this value by 4 to account for uneven distribution of extraterrestrial material on the Earth's surface.)

Bibron *et al.*¹⁸ measured ⁵³Mn activity in the Antarctic snow. The estimated accretion rate in this study was about 30,000 tons yr⁻¹ assuming a uniform flux, but was later multiplied by a factor of 3 to approximate the global accretion rate.

Although our reported flux value can be compared with those from other studies of polar regions, comparison with accretion rate values from oceanic sediments is difficult. Because of the short sampling period in this work (1 yr), extraterrestrial particles collected on our filters are from the sporadic flux of material from solar dust cloud. Particles that make up this background flux are believed to have masses between 10⁻¹³ and 10⁻⁶ g (ref. 22). On the other hand, due to very slow accumulation rate, sediment samples used to estimate the flux of extraterrestrial material usually cover a time span of several million years. Samples that integrate such large time intervals, in addition to background dust particles, also include fragmentation products from large bodies. Extraterrestrial objects larger than 10⁶ g do not frequently enter the earth's atmosphere, the impact probability decreasing with increasing radius; however, over several million years, the impact of objects as large as 10¹⁷ g has a finite probability²³. As the chemical methods used do not discriminate between background dust particles and fragmentation products from large bodies, a significant fraction of the flux estimated from sediment samples may be due to such large body impacts²³. Satellite, visual and radio-meteor studies give a flux of background dust particles of around 16,000 tons yr⁻¹ (refs 22, 24).

Our estimate of 6,000 to 11,000 tons yr⁻¹ for the particles smaller 10⁶ g is in reasonable agreement with these studies.

Due to relatively fast snow accumulation rate in the polar regions (>10 cm yr⁻¹), the time interval covered by each snow and ice sample rarely exceeds 100 years. Since the contribution of large bodies is not significant for short time intervals, the accretion rate estimates from polar samples should be smaller than similar estimates from oceanic sediments.

We thank the NBS reactor personnel, especially N. Bickford for cooperation during irradiations, N. K. Aras for helpful discussions during the development of the radiochemical separation method, and G. E. Gordon for careful review of the manuscript. Supported by an NSF grant.

Received 7 January; accepted 9 July 1987.

- Lal, D., Peters, B. in *Handbuch Der Physik* Vol. 46(2) 551–612 (ed. Flugge, S.) (Springer, Berlin, 1967).
- Cunningham, W. C. & Zoller, W. H. *J. Aerosol Sci.* **12**, 367–384 (1981).
- Maenhaut, W., Zoller, W. H. & Coles, D. G. *J. geophys. Res.* **84**, 2421–2431 (1979).
- Zoller, W. H., Parrington, J. R., Kotra, J. M. P. *Science* **222**, 1118–1121 (1983).
- Tuncel, G. thesis, Univ. Maryland (1985).
- Parkin, D. W. & Tilles, D. *Science* **159**, 936–946 (1968).
- Fenner, F. D. & Presley, B. J. *Nature* **312**, 260–263 (1984).
- Taylor, S. R. *Geochim. cosmochim. Acta* **28**, 1273–1285 (1972).
- Crockett, J. H. & Kuo, H. Y. *Geochim. cosmochim. Acta* **43**, 831–842 (1979).
- Boutron, C. & Lorius, C. *Nature* **277**, 1–3 (1979).
- Legrand, M. R. & Delmas, R. J. *Atmos. Environ.* **18**, 1867–1874 (1984).
- Kumai, M. J. *atmos. Sci.* **33**, 833–841 (1976).
- Turco, R. P., Toon, O. B., Hammil, P. & Whitten, R. C. *J. geophys. Res.* **86**, 1113–1128 (1981).
- Sulek, A., M.S. thesis, Univ. Maryland (1979).
- Thompson, L. G. & Thompson, E. M. *Science* **212**, 812–814 (1981).
- Mason, B. *US geol. Surv. Prof. Pap.* **440**–8–1, (1979).
- Brocas, J. & Piciotto, E. J. *geophys. Res.* **72**, 2229–2236 (1967).
- Bibron, R. R. *et al. Earth planet. Sci.* **21**, 109–116 (1974).
- Ganapathy, R. *Science* **220**, 1158–1161 (1983).
- Uematsu, M. & Duce, R. A. *Nature* (in the press).
- Takahashi, H., Yokoyama, Y., Fireman, E. L. & Lorius, C. *Lunar planet. Sci.* **9**, 1131–1133 (1979).
- Hughes, D. W. in *Cosmic Dust* (ed. McDonnell, J. A. M.) 123 (Wiley, Chichester, 1978).
- Kyte, F. T. & Wasson, J. T. *Science* **232**, 1225 (1986).
- Grun, E., Zook, H. A., Fechtig, H. & Giese, R. H. *Icarus* **62**, 244 (1982).
- Murrell, M. T., Davis, P. A., Nishisumi, K. & Millard, H. T. *Geochim. cosmochim. Acta* **44**, 2067–2074 (1980).
- McGorkell, R., Fireman, E. L. & Langway, C. C. *Science* **156**, 1690–1692 (1967).
- Mulotak, V. I. *Kosmochim. Metroid. 6th Mater. Vses. Symp.* 64–72 (1982).
- Shedlovsky, J. P. & Paisley, S. *Tellus* **18**, 499–503 (1966).
- Patterson, H. & Rostchi, H. *Geochim. cosmochim. Acta* **2**, 81–90 (1952).
- Bonner, F. T. & Laurence, A. S. *Nature* **207**, 933–935 (1965).
- Harris, R. C., Crockett, J. H. & Stainton, M. *Geochim. cosmochim. Acta* **32**, 1049–1056 (1968).
- Barker, J. L. & Anders, E. *Geochim. cosmochim. Acta* **32**, 627–645 (1968).
- Ozima, M., Tagayanagi, M. & Zashu, S. *Nature* **311**, 448–450 (1984).

Models and observations of the impact of natural hydrocarbons on rural ozone

M. Trainer*, E. J. Williams*, D. D. Parrish*†, M. P. Buhr*, E. J. Allwine‡, H. H. Westberg‡, F. C. Fehsenfeld & S. C. Liu

Aeronomy Laboratory, Environmental Research Laboratories, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

* Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

† Department of Chemistry, Metropolitan State College, Denver, Colorado 80204, USA

‡ Laboratory for Atmospheric Research, Washington State University, Pullman, Washington 99164, USA

Large quantities of non-methane hydrocarbons (NMHCs) are emitted into the atmosphere from vegetation^{1,2}. A recent inventory by Lamb *et al.*² indicates that the emission of natural hydrocarbons is significant compared to that of anthropogenic NMHCs in most regions of the United States. Because of their chemical activity, the natural NMHCs can play important parts in the formation of trace gases, such as ozone (O₃), peroxyacetyl nitrate (PAN) and oxygenated secondary hydrocarbons^{3–5}, which contribute to regional-scale air pollution⁶ and may be harmful to crops and

forest⁷⁻⁹. The impact of natural hydrocarbons on the formation of O_3 in the atmosphere has been discussed previously⁴⁻¹³. In general, it has been concluded that their impact is small¹⁴⁻¹⁶. But lack of data on the ambient concentrations of key photochemical species and an incomplete analysis of the photochemistry has prevented a definitive evaluation of the impact of natural NMHCs on rural O_3 . Here we report on concentrations of key trace gases measured concurrently at a rural site in the eastern USA during the summer of 1986 and a modelling study conducted to analyse these measurements. This study demonstrates that natural NMHCs can have a significant impact on ozone formation in rural air.

The field measurements were carried out at Scotia, Pennsylvania. The site is located in a forest preserve 10 km west of State College, remote from the influence of direct anthropogenic emissions. The site is a monitoring station of the National Acid Precipitation Assessment Program (NAPAP) with well-characterized meteorology. Trace gases measured at the site include NO , NO_2 , HNO_3 , PAN, NO_y (the sum of the reactive nitrogen species¹⁷), O_3 , NO_3^- , SO_4^{2-} , individual HC up to C_{10} including the terpenes and isoprene, CH_2O and CH_3CHO .

Ozone shows a characteristic diurnal cycle at the site (Fig. 1). At night O_3 is suppressed due to surface deposition in a shallow inversion layer. In the morning hours O_3 increases following the breakup of the nocturnal inversion layer as air from above is mixed downward. During the day O_3 also increases due to photochemical formation. The highest O_3 concentrations at the site are observed under high-pressure conditions associated with low surface winds, high temperatures and predominantly clear skies. These high-pressure systems can persist for several days, allowing the accumulation of photochemically active trace compounds and the formation of secondary products. Accordingly, high- O_3 concentrations are frequently observed over large regions during these periods^{18,19}.

In particular, during 4-7 July 1986, central Pennsylvania was at the northern edge of a stagnant high-pressure system over the eastern USA. The O_3 concentration observed at Scotia during this 4-day period peaked at 110 p.p.b.v. Figure 2 shows the daytime average hydrocarbon mixing ratios observed at the site; the natural NMHC isoprene was the predominant hydrocarbon. Isoprene concentrations were low in the morning hours and increased throughout the day as the temperature increased, consistent with the expected temperature dependence of the emission rate²⁰. Furthermore, the average isoprene concentration increased during the 4-day period as did the average daytime temperature. The anthropogenic NMHCs (sum of C_3 - C_{10} HC) observed during this period were mainly alkanes, with propane, butanes and pentanes contributing 60% by mass to the C_3 - C_{10} HC. The concentrations of more reactive anthropogenic NMHCs were low. For example, the concentration of ethene was $0.23 \mu g m^{-3}$, while the propene concentration was $<0.1 \mu g m^{-3}$. The high contribution of the less reactive light alkanes to the total burden of anthropogenic NMHCs is characteristic for a rural site²¹.

To simulate this event a fine-resolution, one-dimensional planetary boundary layer (PBL) model developed by Trainer *et al.*²² that includes isoprene photochemistry was used to evaluate the effect of natural NMHCs on the formation of rural ozone. The fine-height resolution is needed to model this event. Because the vertical mixing near the surface is slow and the photochemical destruction of the very reactive natural NMHCs is fast, the mixing ratios of the natural NMHCs decrease rapidly with altitude²³.

In the model calculations the temperature dependence of the isoprene emission rate was described according to measurements by Lamb *et al.*²⁰ at a forest site in eastern Pennsylvania. For the height at which the samples were taken (5 m) the maximum isoprene concentration predicted by the model is 5 p.p.b.v. or $15 \mu g m^{-3}$, which is typical of the observed values.

The emission rate for NO_x was chosen to reproduce the

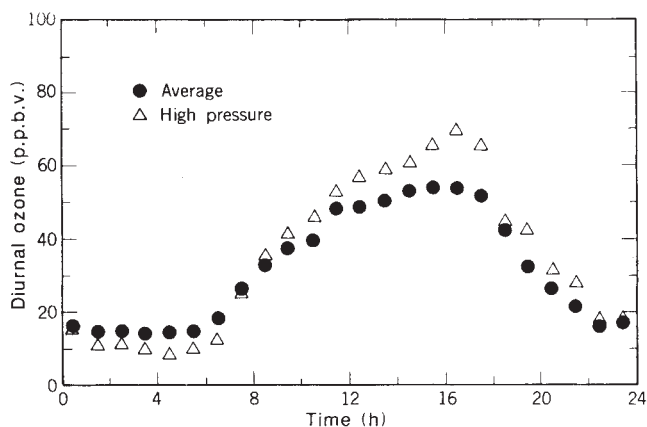


Fig. 1 Average diurnal ozone mixing ratio at Scotia, Pennsylvania, for 25 June-26 July 1986 (circles) and for times when the surface pressure was larger than the average pressure of 970 mbar (open triangles). The sampling height was 5 m above the ground.

observed diurnal pattern of NO_x . During the 4-day period, NO_x showed a sharp peak in the morning that lasted 1-2 h and coincided with the breakup of the nocturnal inversion layer (Fig. 3). The NO_x concentrations during the day were ~ 1 p.p.b.v. As there are no appreciable local NO_x sources, the NO_x is probably transported to the site from distant sources. To simulate this phenomena a continuous NO_x emission rate was assumed for the height range from 200 to 800 m. The exact range is not important provided that it is above the nocturnal inversion layer. During the night-time the vertical mixing is slow and NO_x accumulates in this emission layer. The surface, where the measurements are taken, is isolated from the air aloft during the night by the shallow inversion layer. In the morning, when the inversion breaks up, downward mixing of NO_x leads to the sharp early morning rise in the near-surface NO_x mixing ratio. The subsequent decline in NO_x is due to dilution as the PBL height increases during the daytime and due to photochemical conversion of NO_x to PAN and HNO_3 .

The rate of the NO_x emission, chosen to reproduce the observed diurnal variation, is $4 \times 10^{15} m^{-2} s^{-1}$. This value agrees with the average NO_x emission rate estimated for western Pennsylvania²⁴. The emission rates of the anthropogenic NMHCs were chosen to reproduce the concentrations observed at the site. The ratio of the anthropogenic NMHCs to NO_x emission rates (by volume) was 1.2 with the following partitioning: C_2H_6 , 38%; C_3H_8 , 25%; C_2H_4 , 6%; C_3H_6 , 6%. Butane was taken as the surrogate for the C_4 - C_{10} hydrocarbons.

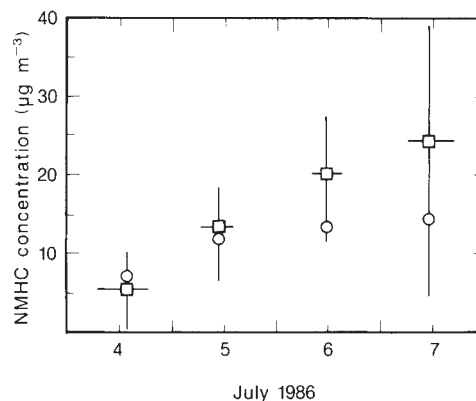


Fig. 2 Average isoprene (squares) and anthropogenic C_3 - C_{10} HC (circles) mixing ratios at Scotia, for 4-7 July 1986. Horizontal bars, times during which several individual samples were taken; vertical bars, variation of isoprene throughout the time.

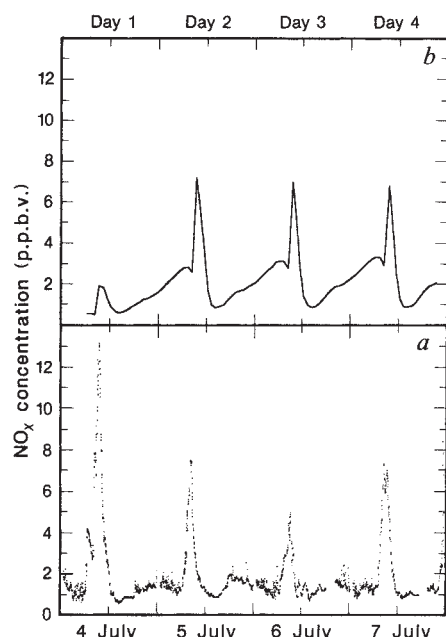


Fig. 3 *a*, Observed variation of the NO_x mixing ratio at Scotia for 4–7 July, 1986; *b*, the variation of the NO_x mixing ratio during a 4-day period as calculated by the model for the emission model described in the text.

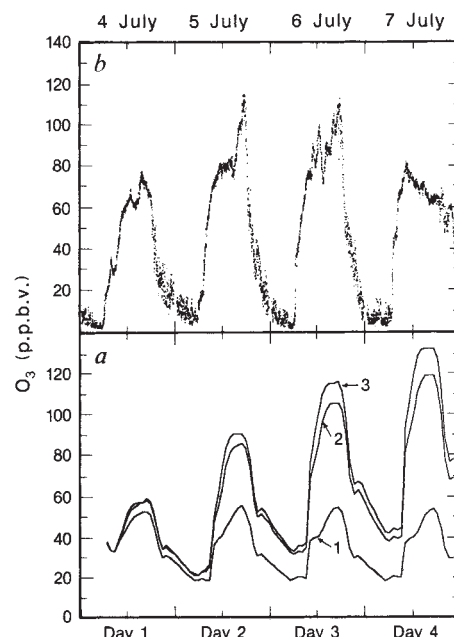


Fig. 4 *a*, Model-calculated variation of the O_3 mixing ratio during a 4-day period for the following models: 1, CO , CH_4 oxidation only; 2, additional isoprene emission; 3, additional isoprene and anthropogenic NMHC emission. *b*, Observed variation of the O_3 mixing ratio at Scotia, 4–7 July 1986.

Figure 4a shows the predicted ozone variation for a 4-day period at a height of 5 m for the NO_x emission rate and three models: only CO and CH_4 , with isoprene, and with anthropogenic NMHCs. The model that includes all the hydrocarbon emissions simulates the observed peak ozone concentrations well except for 7 July. The decline in O_3 observed on 7 July may indicate an approaching cold front and change of air masses at the site. This effect needs to be investigated in more detail. If no hydrocarbons are emitted and only CO and CH_4 oxidation are considered in the model (Fig. 4a, curve 1), the daytime maximum ozone values are ~ 55 p.p.b.v. When isoprene emission is added (Fig. 4a, curve 2), the concentrations of HO_2 and organic peroxy radicals are substantially increased, resulting in a strong build-up of ozone during the 4-day period. In this context we note that carbonyl compounds such as methylvinylketone, methacrolein and formaldehyde are major intermediates in the production of HO_2 and organic peroxy radicals from isoprene. These intermediates and other byproducts such as PAN, H_2O_2 and organic peroxides are increased substantially due to the oxidation of isoprene²². If there is an efficient heterogeneous removal of these intermediates, the build-up of peroxy radicals and thus ozone would be significantly curtailed.

Additional inclusion of the emission of anthropogenic hydrocarbons (Fig. 4a, curve 3) shows only a small additional increase in the ozone production in these conditions. The system is in a strongly nonlinear regime where an increase of the hydrocarbon emission rate does not lead to a corresponding increase in the ozone production rate. Consequently, the model evaluation of the impact of natural NMHCs depends on the way the question is posed. For example, in a calculation that includes first the emission of anthropogenic NMHCs and then the addition of isoprene, most of the ozone increase ($\sim 70\%$) is due to anthropogenic NMHCs. Many previous studies that considered only this model may have therefore underestimated the contribution of natural NMHCs to rural ozone formation.

The comparison of model prediction and observation indicates that the oxidation of NO_x , which is mainly of anthropogenic origin, in the presence of naturally emitted isoprene, at levels observed at this rural site, can lead to high ozone concentrations during a stagnant high-pressure period.

If high ozone in rural areas is of concern, a reduction of the emission of anthropogenic hydrocarbons will probably not reduce the ozone formation substantially in rural air. Consideration has to be given to a reduction of the anthropogenic NO_x emissions.

Finally it is interesting to note that, in the absence of anthropogenic NO_x emissions, isoprene contributes little to ozone formation. Model calculations that assume only a biogenic surface emission of NO_x with a flux of $2 \times 10^{14} \text{ m}^{-2} \text{ s}^{-1}$ (ref. 25) indicate that isoprene has little effect on the production of ozone and even can lead to a slight reduction in O_3 . The conversion of NO_x to PAN and possibly other organic nitrates in the photo-oxidation of isoprene plays a critical part in moderating the PBL photochemistry²⁶.

This research has been funded as part of NAPAP by the National Oceanic and Atmospheric Administration.

Received 2 April; accepted 4 September 1987.

1. Zimmerman, P. R. *EPA Rep. No. 904/9-77-028* (US EPA, 1979).
2. Lamb, B., Guenther, A., Gay, D. & Westberg, H. *Atmos. Environ.* **21**, 1695–1705 (1987).
3. Demerjian, K. L., Kerr, J. A. & Calvert, J. G. *Adv. envir. Sci. Technol.* **4**, 1–262 (1974).
4. Graedel, T. E. *Rev. Geophys. Space Phys.* **17**, 937–946 (1979).
5. Dimitriadis, B. *J. Air Pollut. Control. Ass.* **31**, 229–235 (1981).
6. Environmental Protection Agency *Investigation of Rural Oxidant Levels as related to Urban Hydrocarbon Control Strategies Rep. No. 450/3-75-036* (1975).
7. Heck, W. W. *et al. J. Air Pollut. Control Ass.* **34**, 729–735 (1984).
8. Wang, D., Bormann, F. H. & Karnosky, D. F. *Envir. Sci. Technol.* **20**, 1122–1125 (1986).
9. Skärby, L. & Sellden, G. *Ambio* **13**, 68–72 (1984).
10. Lonneman, W. A., Seila, R. L. & Bufalini, J. J. *Envir. Sci. Technol.* **12**, 459–463 (1978).
11. Sandberg, J. S., Basso, M. J. & Okin, B. A. *Science* **200**, 1051–1054 (1978).
12. Arnst, R. R. & Meeks, S. A. *Atmos. Environ.* **15**, 1643–1651 (1981).
13. Isaksen, I. A., Hov, Ø & Hessvedt, E. *Envir. Sci. Technol.* **12**, 1279–1284 (1978).
14. Lurmann, F. W., Lloyd, A. C. & Nitta, B. *Atmos. Environ.* **17**, 1951–1963 (1983).
15. Lurmann, F. W., Nitta, B., Ganesan, K. & Lloyd, A. C. *Atmos. Environ.* **18**, 1133–1143 (1984).
16. Altshuller, A. P. *Atmos. Environ.* **17**, 2131–2165 (1983).
17. Fahey, D. W. *et al. J. geophys. Res.* **91**, 9781–9793 (1986).
18. Vukovich, F. M., Bach, W. D. Jr, Grissman, B. W. & King, W. J. *Atmos. Environ.* **11**, 967–984 (1977).
19. Guichert, R. & van Dop, H. *Atmos. Environ.* **11**, 145–155 (1977).
20. Lamb, B., Westberg, H., Allwine, G. & Quarles, T. J. *geophys. Res.* **90**, 2380–2390 (1985).
21. Sexton, K. & Westberg, H. *Atmos. Environ.* **18**, 1125–1132 (1984).
22. Trainer, M. *et al. J. geophys. Res.* (in the press).
23. Hov, Ø, Schjoldager, J. & Wathne, B. M. J. *geophys. Res.* **88**, 10679–10688 (1983).
24. National Acid Precipitation Assessment Program *Second Annual Acid Deposition Emission Inventory Symposium EPA Rep. No. 600/9-86-010* (1986).
25. Williams, E. J., Parrish, D. D., Buhr, M. P., Fehsenfeld, F. C. & Fall, R. *Eos Trans. AGU* **67**, 892 (1986).
26. Jacob, D. J. & Wofsy, S. C. *J. geophys. Res.* (in the press).

Depletion of ^{13}C in lignin and its implications for stable carbon isotope studies

Ronald Benner*, Marilyn L. Fogel†, E. Kent Sprague‡ & Robert E. Hodson§

* Marine Science Institute, University of Texas at Austin, Port Aransas, Texas 78373, USA

† Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC 20008, USA

‡ Center for Applied Isotope Studies, University of Georgia, Athens, Georgia 30604, USA

§ Department of Microbiology, University of Georgia, Athens, Georgia 30602, USA

Stable carbon isotope compositions of organic matter are now widely used to trace carbon flow in ecosystems, and have been instrumental in shaping current perceptions of the importance of terrestrial vegetation to estuarine and coastal marine environments. A general assumption in these and other studies relying on carbon isotope compositions for source identification of organic matter has been that the major biochemical components of plant tissues are isotopically invariant. We report here large differences between the carbon isotope compositions of the polysaccharide and lignin components of a variety of vascular plants, including the salt-marsh grass *Spartina alterniflora*, and demonstrate that the carbon isotope composition of *Spartina* detritus gradually changes during biogeochemical processing as polysaccharides are preferentially removed, leaving a material that is relatively enriched in lignin-derived carbon and depleted in ^{13}C .

Differences among the carbon isotope compositions of the major chemical components of plant tissues will influence interpretations of the sources of organic carbon to food webs and to pools of dissolved and particulate organic matter if selective assimilation and degradation occurs. As there has not been a comprehensive study of the variability in the carbon isotope composition of the major components of vascular plant tissues, we undertook a survey of the isotopic variability in plant components from a wide variety of plant tissues. The structural polysaccharides, cellulose and hemicellulose, account for 57–77% of herbaceous and woody plant tissues (see Table 1). Both cellulose and hemicellulose were typically enriched in ^{13}C by 1–2‰ relative to whole-plant material (Table 1). Lignin occurs

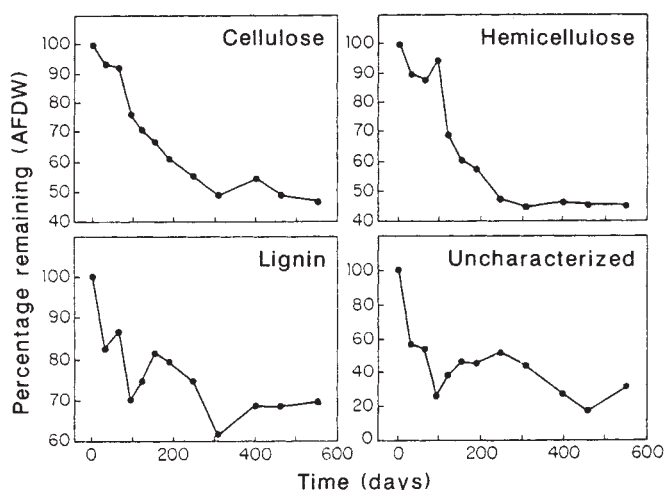


Fig. 1 Percentage of the initial content of each component remaining in *Spartina alterniflora* rhizome material after up to 18 months of below-ground decomposition. Each point is the mean of two replicates. Wet-chemical and gravimetric analyses^{21,22} were used to quantify the cellulose, hemicellulose and uncharacterized fractions. Lignin concentrations, as determined by gas capillary chromatography of lignin-derived phenols released upon oxidation of samples with CuO^{23} , were corrected for yield¹, and the total lignin concentration was calculated as: $(3.3 \times \text{sum of vanillyl phenols}) + (1.1 \times \text{sum of syringyl phenols}) + \text{sum of } p\text{-hydroxy phenols} + \text{sum of cinnamyl phenols}$. The parameters for the regression $\ln X_t = \ln X_0 - kt$ are: cellulose, $k = -0.0014$ ($P < 0.001$), $t_{1/2} = 1.4$ yr; hemicellulose, $k = -0.0016$ ($P < 0.001$), $t_{1/2} = 1.2$ yr; lignin, $k = -0.0005$ ($P < 0.01$), $t_{1/2} = 3.6$ yr; uncharacterized, $k = -0.0017$ ($P < 0.01$), $t_{1/2} = 1.1$ yr, where X is the amount remaining after time t (days); k , the decay constant, is the slope of the regression line; and $t_{1/2}$ is the half-life (yr).

as an integral cell-wall component of all vascular plants and is not found in algae or other plants that lack vascular tissues¹. Lignin accounts for 17–31% of the biomass of woody tissues and 4–9% of the biomass of herbaceous tissues (see Table 1), but contains ~50% more carbon per unit weight than cellulose and thus accounts for a greater percentage of the total carbon than indicated by a dry weight percentage.

In all plant tissues examined, lignin was depleted in ^{13}C by 2–6‰ relative to whole-plant material and by 4–7‰ relative to cellulose. The aromatic amino acids phenylalanine and tyrosine

Table 1 Stable carbon isotope compositions and ash-free dry weight (AFDW) percentages of the major chemical components of vascular plant tissues

Plant species, photosynthetic mode and tissue analysed		Whole tissue $\delta^{13}\text{C}$	Hemicellulose AFDW $\delta^{13}\text{C}$	Cellulose AFDW $\delta^{13}\text{C}$	Lignin AFDW $\delta^{13}\text{C}$	Uncharacterized AFDW $\delta^{13}\text{C}$
<i>Spartina alterniflora</i> (tall form)	C ₄ leaf	-13.1	41.1 -11.6	30.6 -11.8	6.1 -17.4	22.2 -16.5
<i>Spartina alterniflora</i> (short form)	C ₄ leaf	-12.6	43.6 -11.2	30.8 -11.1	4.3 -17.0	21.3 -16.8
	rhizome	-13.2	37.0 -11.4	30.6 -12.3	9.3 -17.9	23.1 -15.4
<i>Cynodon dactylon</i>	C ₄ leaf	-18.7	32.8 -14.5	25.5 -18.2	7.9 -25.4	33.8 -24.2
<i>Juncus roemerianus</i>	C ₃ leaf	-23.6	46.3 -22.5	28.3 -22.5	6.0 -26.8	19.4 -26.8
	rhizome	-23.7	36.7 -22.9	21.0 -23.0	5.1 -27.3	37.2 -24.4
<i>Carex walteriana</i>	C ₃ leaf	-25.1	38.3 -24.0	39.1 -24.4	4.4 -29.3	18.2 -27.9
<i>Pinus elliotii</i>	C ₃ wood	-25.3	25.4 -24.3	36.3 -23.8	31.0 -27.9	7.3 -25.2
<i>Juniperus virginiana</i>	C ₃ wood	-24.0	23.0 -22.5	34.4 -22.1	31.4 -26.5	11.2 -25.9
<i>Quercus niger</i>	C ₃ wood	-27.5	26.5 -28.4	41.6 -27.2	20.8 -30.8	11.1 -20.3
<i>Acer rubrum</i>	C ₃ wood	-27.2	31.1 -27.6	44.4 -25.9	17.2 -29.2	7.3 -28.7

Values of $\delta^{13}\text{C}$ (per mil) relative to that for the PDB carbonate standard are defined by the equation $\delta^{13}\text{C} = [({}^{13}\text{C}/{}^{12}\text{C}_{\text{sample}}/{}^{13}\text{C}/{}^{12}\text{C}_{\text{standard}}) - 1] \times 10^3$. Sequential detergent analyses^{21,22} were used to isolate four chemical fractions: hemicellulose–cellulose–lignin, cellulose–lignin, cellulose and lignin. Weight percentages and stable carbon isotope compositions ($\pm 0.1\%$) of cellulose and lignin were determined directly, whereas these values were calculated by mass balance for the hemicellulose and uncharacterized fractions. The $\delta^{13}\text{C}$ of purified cellulose (Sigma) did not change (to $\pm 0.1\%$) after treatment with the various detergent solutions, indicating that the detergents were completely removed during washing and did not interfere with the determination of $\delta^{13}\text{C}$ values.

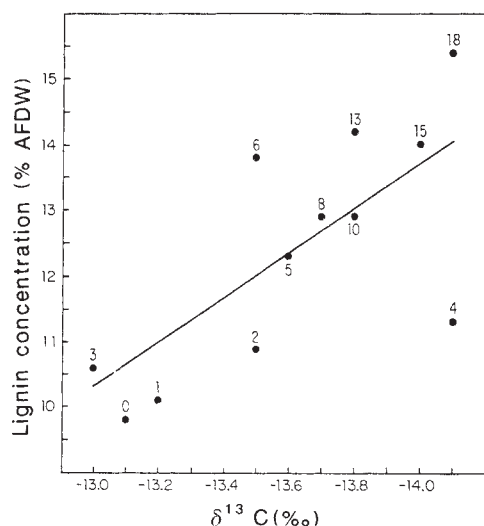


Fig. 2 Stable carbon isotope compositions and lignin concentrations in *Spartina alterniflora* rhizome material that was buried at depths of 5–10 cm in salt-marsh sediments and retrieved at 1–3-month intervals over an 18-month period. Integers above points represent the number of months of burial. Lignin was quantified as described in Fig. 1 legend. A linear regression of the data is shown ($r = -0.73$, $P < 0.01$). Replicate analyses of independent samples ($n = 2$) or replicate analyses of the same sample were made with maximum standard deviations of $\pm 1.4\%$ for lignin concentrations and $\pm 0.2\%$ for stable carbon isotope compositions.

are biosynthetic precursors of lignin¹, and it appears that most of the fractionation against ^{13}C which can be measured in lignin occurs during biosynthesis of phenylalanine and tyrosine. There are large variations in $\delta^{13}\text{C}$ values among amino acids², and tyrosine, in particular, can be relatively depleted in ^{13}C (ref. 3). Tyrosine is an important precursor of lignin in grasses, but is not an efficient precursor of lignin in conifers and woody angiosperms¹. The relative contributions of phenylalanine and tyrosine to lignin biosynthesis may influence the overall depletion of ^{13}C in lignin and thereby account for the greater depletion of ^{13}C observed in grass lignins relative to wood lignins (see Table 1). Hemicellulose, cellulose and lignin together account for 70–90% of most plant tissues. We did not chemically characterize the remaining 10–30% of plant biomass, which contains a variety of compounds including lipids, carbohydrates and amino acids. This chemically uncharacterized fraction was typically depleted in ^{13}C by 1–4% relative to whole-plant material (Table 1).

The different chemical components of plant tissues vary considerably in their susceptibility to microbial degradation and assimilation by animals. Numerous studies have demonstrated that under aerobic and anaerobic conditions, the polysaccharide components of vascular plants, such as *Spartina*, are degraded two to five times more quickly than the lignin component, which results in the gradual enrichment of plant detritus in lignin-derived carbon^{4–8}. As the relative concentration of lignin in the remaining material increases, it follows that the $\delta^{13}\text{C}$ value will decrease from that of the original plant material to values more similar to that of the lignin component. A study of the below-ground (5–10-cm depth) decomposition of *Spartina* rhizome material in salt-marsh sediments was initiated to determine if, in natural conditions, the $\delta^{13}\text{C}$ of this material decreased as the relative lignin content increased. The concentrations of cellulose, hemicellulose, lignin and the uncharacterized fraction were monitored during the decomposition study, and Fig. 1 shows the percentage remaining of the original amount of each of these components at various times. The exponential decay function was used to model the loss of each component, and in all cases the model gave a statistically significant fit to the data. The decay constant (k) and half-life ($t_{1/2}$) of each component were

Table 2 Measurements and model predictions of total percentage of material remaining, lignin concentration and stable carbon isotope composition for decaying *Spartina alterniflora* rhizome material in salt-marsh sediments

Decomposition period (yr)	Material remaining (% AFDW)	Lignin concentration (% AFDW)	$\delta^{13}\text{C}$ (‰)
0.0 measured	100.0	9.8	-13.1
1.5 measured	44.9	15.4	-14.1
1.5 predicted	45.6	16.5	-14.0
5.0 predicted	9.2	42.1	-15.4
10.0 predicted	1.8	81.4	-17.2

Data using kinetic parameters determined from the below-ground decomposition study values were calculated using a multiple exponential decay model⁹, $X_t = C_0 e^{(-k_c t)} + H_0 e^{(-k_h t)} + L_0 e^{(-k_l t)} + U_0 e^{(-k_u t)}$, where X is the amount remaining after time t (days), C_0 is the initial cellulose concentration (29.5%), H_0 is the initial hemicellulose concentration (32.4%), L_0 is the initial lignin concentration (9.8%), U_0 is the initial concentration of the chemically uncharacterized fraction (28.3%) and k is the decay constant (see Fig. 1 legend for decay constants). Weight percentages of lignin remaining were carbon-normalized ($1.5 \times \text{wt } \%$) with respect to polysaccharides in the calculation of predicted $\delta^{13}\text{C}$ values. Model predictions of $\delta^{13}\text{C}$ values were calculated using the following mass balance equation: $\delta^{13}\text{C}_{\text{whole tissue}} = (\text{cellulose fraction}) (-12.3\%) + (\text{hemicellulose fraction}) (-11.4\%) + (\text{lignin fraction}) (-17.9\%) + (\text{uncharacterized fraction}) (-15.4\%)$. Values for stable carbon isotope ratios of each component of *Spartina* rhizome material are taken from Table 1.

determined from this function (see Fig. 1 legend). Lignin was the most resistant component to degradation; the half-life for *Spartina* lignin in these predominantly anaerobic sediments was 3.6 yr. The relative enrichment in lignin-derived carbon was accompanied by a marked decrease in the $\delta^{13}\text{C}$ of the remaining detritus (Fig. 2).

Implicit in the use of the exponential decay function to model decomposition is the assumption that all constituents of the substrate have the same reactivity. Heterogeneous polymers, such as lignin, or mixtures of different compounds, such as occur in the uncharacterized fraction of *Spartina*, contain components that vary in their susceptibility to decomposition and may therefore not have a well defined half-life. Recognizing the difficulties in modelling the fate of complex organics, we present the following model to demonstrate the general features of the decomposition of *Spartina* detritus, rather than to predict an absolute timescale.

The decay constants derived for each of the major fractions of *Spartina* were used in a multiple exponential decay model⁹ to predict the long-term fate of these biopolymers in salt-marsh sediments and the $\delta^{13}\text{C}$ value of the remaining detritus (see Table 2 legend). The $\delta^{13}\text{C}$ value of -14.0‰ predicted by the model is within experimental error of the $\delta^{13}\text{C}$ value (-14.1‰) measured at the end of the 1.5-yr field study (Table 2). Thus, it appears that the $\delta^{13}\text{C}$ values of the individual biochemical components of *Spartina* do not change significantly during decomposition and that the observed decreases in $\delta^{13}\text{C}$ are predictable using a mass-balance approach. The model predicts that after 10 years of decomposition the remaining *Spartina* detritus is mostly derived from lignin and has a stable carbon isotope composition of -17.2%. Previous investigators have observed^{10–12} that the $\delta^{13}\text{C}$ of *Spartina* detritus collected from estuarine sediments and waters is often more negative (-15 to -19‰) than values for fresh *Spartina*, but a mechanism for these changes in isotopic signature has not previously been demonstrated. The data presented here indicate that the change in isotopic composition is due to the relative resistance of the isotopically light lignin fraction to microbial degradation.

Changes in carbon isotope composition during the decomposition of plant matter have, in general, been assumed to be

minimal. The decomposition model presented here for *Spartina* detritus indicates, however, that relatively large (4‰) isotope changes can occur during the early diagenesis of lignified plant tissues and that shifts in $\delta^{13}\text{C}$ should be considered in source identification of organic matter. Indeed, recent analyses of the carbon isotope and chemical composition of ancient buried woods have shown that the remaining tissues are relatively depleted in ^{13}C and enriched in lignin relative to modern counterparts¹³.

The relative contributions of *Spartina alterniflora* grass (whole-tissue $\delta^{13}\text{C} = -12$ to -13%) and phytoplankton ($\delta^{13}\text{C} = -19$ to -23%) to estuarine organic matter have been debated for many years¹⁴⁻¹⁶. Early studies of fluxes of organic matter in salt-marsh estuaries indicated that a large portion of the detritus derived from *Spartina alterniflora* was transported to adjacent waters and sediments^{14,15}. Data from more recent studies based on the carbon isotope compositions of estuarine detritus were interpreted to indicate that *Spartina* detritus was relatively unimportant to carbon flow beyond marsh sediments¹⁷⁻²⁰. The simple mixing models typically used to distinguish different carbon sources on the basis of $\delta^{13}\text{C}$ values do not take into account shifts in isotope composition during decomposition and assimilation and, in the case of *Spartina*, may have significantly underestimated the contribution of salt-marsh vegetation to estuarine and coastal waters.

We thank J. I. Hedges, R. D. Guy, W. J. Wiebe and W. B. Whitman for comments on an earlier version of the manuscript.

This research was supported by NSF and the NOAA Office of Sea Grants (to R.E.H. and R.B.). Additional funding (to E.K.S.) was provided by the American Museum of Natural History.

Received 12 June; accepted 13 August 1987.

1. Sarkanen, K. V. & Ludwig, C. H. (eds) *Lignins: Occurrence, Formation, Structure and Reactions* (Wiley, New York, 1971).
2. Abelson, P. H. & Hoering, T. C. *Proc. natn. Acad. Sci. U.S.A.* **47**, 623-632 (1961).
3. Macko, S. A., Fogel, M. L., Hare, P. E. & Hoering, T. C. *Chem. Geol.* **65**, 79-92 (1987).
4. Maccubbin, A. E. & Hodson, R. E. *Appl. envir. Microbiol.* **40**, 735-740 (1980).
5. Benner, R., Maccubbin, A. E. & Hodson, R. E. *Appl. envir. Microbiol.* **47**, 998-1004 (1984).
6. Wilson, J. O. *Appl. envir. Microbiol.* **49**, 478-484 (1985).
7. Benner, R., Maccubbin, A. E. & Hodson, R. E. *Microb. Ecol.* **12**, 291-298 (1986).
8. Benner, R., Moran, M. A. & Hodson, R. E. *Limnol. Oceanogr.* **31**, 89-100 (1986).
9. Westrich, J. T. & Berner, R. A. *Limnol. Oceanogr.* **29**, 236-249 (1984).
10. Schwinghamer, P., Tan, F. C. & Gordon Jr. D. C. *Can. J. Fish. aquat. Sci.* **40**, 262-272 (1983).
11. Ember, L. M., Williams, D. F. & Morris, J. T. *Mar. Ecol. Prog. Ser.* **36**, 33-42 (1987).
12. Sprague, E. K., Frey, R. W., Fogel, M. L. & Gize, A. P. *Abstr. a. Meet. Soc. econ. Paleont. Miner.* **3**, 105 (1986).
13. Spiker, E. C. & Hatcher, P. G. *Geochim. cosmochim. Acta* **51**, 1385-1391 (1987).
14. Teal, J. M. *Ecology* **43**, 614-624 (1962).
15. Odum, E. P. & de la Cruz, A. A. in *Estuaries* (ed. Lauff, G. H.) 383-388 (Am. Ass. Adv. Sci., Washington, 1967).
16. Nixon, S. W. in *Estuarine and Wetland Processes* (eds Hamilton, P. & Macdonald, K. B.) 437-525 (Plenum, New York, 1980).
17. Haines, E. B. *Estuar. coast. mar. Sci.* **4**, 609-616 (1976).
18. Haines, E. B. *Oikos* **29**, 254-260 (1977).
19. Haines, E. B. & Montague, C. L. *Ecology* **60**, 48-56 (1979).
20. Sherr, E. B. *Geochim. cosmochim. Acta* **46**, 1227-1232 (1982).
21. Goering, H. K. & Van Soest, P. J. *Agric. Handb.* No. 379, 1-20 (US Department of Agriculture, Washington, 1970).
22. Van Soest, P. J. & Robertson, J. B. in *Standardization of Analytical Methodology for Feeds* (eds Balch, C. C. & Graham, M.) 49-60 (International Development Research Centre (IDRC), Ottawa, 1980).
23. Hedges, J. I. & Ertel, J. R. *Analyt. Chem.* **54**, 174-178 (1982).

Was the Archaean biosphere upside down?

James C. G. Walker

Departments of Atmospheric and Oceanic Sciences and Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

Photosynthesis produces reduced organic carbon and an oxidized partner in equivalent molar amounts. These compounds can react with one another, again in equivalent molar amounts, so that no net change occurs in the overall level of oxidation of the biosphere (here taken to mean the biota together with the part of the Earth with which living things interact¹). But, the reduced and oxidized partners have different susceptibilities to transport by environmental processes and so, typically, they become separated. Conservation of matter implies that for every mole of excess oxidant in an oxidizing region of the biosphere there must be a mole of excess reductant in a reducing region. Today, the oxidized partner in photosynthesis is usually free oxygen, which floats upward to accumulate in excess in the atmosphere, whereas organic matter settles downward to collect in sediments and stagnant pools. On the anoxic Archaean Earth, the oxidized partner was probably iron. As oxidized iron is markedly less soluble and mobile than organic carbon, differential transport in the Archaean biosphere would have had an effect just the opposite of that in the modern biosphere. The oxidized partner would have settled downward more rapidly than the reduced partner, resulting in the accumulation of excess oxidant in sediments and stagnant pools. An equivalent excess of the more volatile reduced compounds would have been left behind in ocean and atmosphere in the form of dissolved organic carbon and gaseous hydrocarbons. On average, therefore, the Archaean biosphere may have been oxidizing at the bottom and reducing on top.

For about the first 2,000 Myr of Earth history the world was anaerobic and inhabited only by prokaryotes². Biogeochemical cycles at this time may have differed markedly from those of the present, and an understanding of these cycles can influence ideas about the evolution of early life and the environmental interpretation of the sedimentary rock record. Our ideas about

biogeochemical cycles, however, are prejudiced by our familiarity with the abundantly oxidizing modern environment with its productive eukaryotic biota. By far the dominant form of autotrophy today is oxygenic photosynthesis, in which green plants or cyanobacteria reduce carbon dioxide to cell material and maintain oxidation balance by releasing free oxygen to the environment. This process is largely balanced on the modern Earth by aerobic respiration, in which heterotrophic organisms oxidize organic compounds using free oxygen as electron acceptor. On the Archaean Earth, the major biogeochemical cycles of production and consumption of organic matter must also have involved reduction and oxidation of carbon, but the environmentally abundant oxidation partner of this carbon cannot have been free oxygen when the world was anaerobic. Even if Archaean autotrophs released free oxygen to the environment, this oxygen must have reacted rapidly with reduced inorganic constituents of ocean and atmosphere to maintain an anaerobic environment. The oxidized products of photosynthesis or of this ambient reaction would have been used by anaerobic heterotrophs as electron acceptors.

The obvious oxidation partners for photosynthetic carbon on the early Earth are sulphur or iron. Iron is, of course, geochemically much more abundant than sulphur. Nearly all sulphur in igneous rocks is in the form of the iron sulphide, pyrite. When pyrite weathers under aerobic conditions today, sulphur enters solution in the oxidized form as sulphate; insoluble iron oxide remains behind as a solid. Sulphur is therefore more abundant in ocean and atmosphere today than iron, but it would not have been more abundant on the anaerobic Earth³. Weathering under anaerobic conditions would not mobilize highly insoluble pyrite, but would allow the iron in abundant iron silicate minerals to enter solution in the soluble ferrous oxidation state. It is therefore likely that dissolved iron was considerably more abundant than dissolved sulphur in the Archaean hydrosphere. Iron could have been the oxidation partner of carbon that permitted productive biogeochemical cycles without free oxygen. For the purposes of my argument it does not matter whether the iron was oxidized directly in the metabolic apparatus of the photosynthetic organisms or indirectly as a result of reactions in the ambient medium between dissolved reduced iron and the oxidized direct product of photosynthesis. What matters is that the

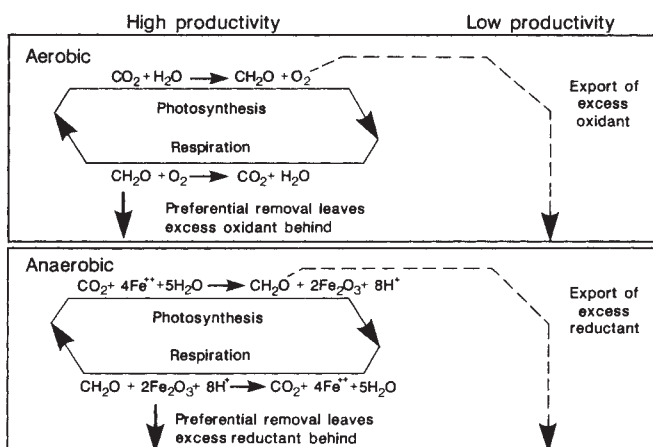


Fig. 1 Schematic view of the biogeochemical cycles of photosynthesis and respiration, contrasting an aerobic biosphere (top) with an anaerobic biosphere (bottom). The chemical reactions indicate the geochemical consequences of these metabolic processes, not necessarily the specific metabolic reactions involved. The symbol CH_2O represents organic matter. Reduction of carbon in the aerobic biosphere is coupled to the release of free oxygen; in the anaerobic biosphere it is coupled to oxidation of iron. Photosynthesis and respiration produce and consume oxidized and reduced compounds in equivalent amounts, so the biogeochemical cycle is nearly closed.

oxidizing partners of the reduced organic compounds produced by photosynthesis appeared, in the environment, largely as oxidized iron.

On thermodynamic grounds the reaction of organic matter with ferric iron is expected to release more energy than reaction with sulphate. Measurement of pore-water composition in anoxic sediments shows that iron is mobilized in diagenesis before sulphate is reduced⁴. In short, there is no reason to suppose that heterotrophs on the early Earth did not breathe iron in preference to sulphate. This suggestion is novel. Although Cloud⁵ suggested that iron might have consumed the oxidizing products of photosynthesis on the anaerobic Earth, he did not point out that oxidized iron could have served as the oxidized partner in a closed biogeochemical cycle by supporting heterotrophic respiration.

Oxygen is more volatile than cell material. On the aerobic Earth there is a tendency for oxygen to rise into the atmosphere while organic matter remains at the ground or in particulate or dissolved form in the hydrosphere. Dissolved organic carbon or reduced biogenic gases are consumed in reactions with the excess oxygen in ocean and atmosphere. Particulate organic carbon tends to be denser than water and therefore to settle more rapidly than oxygen to the bottom of oceans or lakes. There, excess organic matter accumulates in soils or lake or ocean sediments to produce locally anaerobic environments. Because oxidized and reduced compounds are produced and consumed in stoichiometrically equivalent amounts, the key to the development of oxidizing or reducing local environments is differential transport of oxidized and reduced compounds. The modern biosphere is oxidizing on the top and reducing on the bottom.

This partitioning of the biosphere, illustrated schematically in Fig. 1, is a consequence of the physical properties of oxygen and organic matter and is not a consequence of any general property of oxidized and reduced chemical compounds. In particular, oxidized iron is less volatile, less soluble, less mobile and more dense than cell material and organic compounds. It is for these reasons that almost all iron on the modern Earth is in solid or particulate form. The partitioning of the Archaean biosphere should therefore have been just the opposite of that of today, with oxidized iron settling more rapidly than reduced compounds to the bottom of water bodies to produce local

concentrations of oxidized sediments in a generally anaerobic world. The more volatile organic compounds, such as methane, would have diffused upwards into the atmosphere more rapidly than oxidized iron to yield a generally reducing atmosphere. It is in this sense that the Archaean biosphere may have been upside down, oxidizing on the bottom and reducing on the top (see Fig. 1).

This statement should have been true in an average sense, applying to the overall properties of the biosphere without regard for local heterogeneity. Consideration of the consequences of heterogeneity in biological productivity reveals that the Archaean biosphere was probably back to front as well as upside down. Regions of high biological productivity today tend to be underlain by sediments rich in organic carbon. The oxygen released when this organic matter was synthesized has been left behind in ocean and atmosphere, where it has been redistributed around the globe. In biologically unproductive regions of the Earth this extra oxygen, stoichiometrically equivalent to the extra organic matter preserved in the sediments below productive regions, consumes what organic matter there is so that the deserts of the world, both on land and at sea, are underlain by oxidized sediments containing little or no organic matter.

Because of the immobility of oxidized iron, the situation would have been quite the opposite on the anaerobic Earth. There, regions of high biological productivity would have been underlain by sediments rich in more rapidly deposited oxidized iron and deficient in organic matter, whereas the more mobile organic matter would have been left in excess in ocean and atmosphere and redistributed around the globe. In unproductive regions of the globe this excess of reduced organic compounds would have made possible the complete consumption of any oxidized iron produced locally. In due course the excess organic matter would have been incorporated into sediments deficient in oxidized iron that accumulated beneath biologically unproductive regions. Because oxidized and reduced compounds are produced and destroyed in stoichiometrically equivalent amounts, the preservation of oxidant in the sediments of one region must have been balanced by the preservation of an equivalent amount of reductant in the sediments of another region. In short, on an anaerobic world, regions of high biological productivity would have deposited oxidized sediments deficient in organic carbon, and the world's deserts would have deposited reduced sediments containing more organic matter than oxidized iron.

Both types of sedimentary rock are known for the Archaean and early Proterozoic Earth. Clastic deposits containing low concentrations of organic matter and essentially no oxidized iron are common⁶. In my interpretation they would have marked the unproductive regions of the globe. The sedimentary rocks rich in oxidized iron and deficient in organic carbon are the banded iron formations that formed in abundance only on the anaerobic Earth^{7,8}. I follow Cloud⁵ in attributing these deposits to the more productive regions of the Archaean biosphere.

There is no contradiction in the suggestion that banded iron formations, which now contain very little organic matter, were deposited under regions of unusually high biological productivity. High biological productivity yielded abundant insoluble oxidized iron, which settled relatively rapidly into sediments to attack and destroy any organic matter incorporated in those sediments. The product of the oxidation of organic matter in banded iron formations can be detected by isotopic analysis of the carbon incorporated in the carbonate minerals they contain^{9,10}. This carbonate exhibits lower relative amounts of the heavy isotope of carbon than does the oxidized carbon in sea water. Isotopically light carbon is characteristic of the organic matter produced by autotrophic organisms, including photosynthetic organisms. The isotopically light carbonate in banded iron formations is evidence that a significant fraction of this carbonate was produced by diagenetic oxidation of isotopically light organic matter incorporated into the sediments when they

were deposited. Diagenesis has consumed organic carbon and ferric iron, leaving a record in the form of isotopically light carbonate minerals.

I suggest, therefore, that biogeochemical cycles on the anaerobic Earth would have yielded relative vertical and horizontal distributions of reduced and oxidized compounds that are just the opposite of those of the aerobic Earth. These ideas offer an explanation of why banded iron formations are deficient in organic matter; they also suggest that Archaean sediments rich in organic carbon are not indicative of high biological productivity in the overlying water column—quite the reverse. This analysis also carries the surprising implication that photosynthesis on the anaerobic Earth, even oxygenic photosynthesis, could have increased the concentrations of reduced gases in the atmosphere¹¹. The peculiar aspects of an anaerobic biosphere should be considered when the environments of early life are being deduced from the sedimentary rock record.

This paper describes research supported in part by NSF.

Received 20 March; accepted 12 August 1987.

1. Valentine, J. W. *Evolutionary Paleocology of the Marine Biosphere* (Prentice-Hall, Englewood Cliffs, 1973).
2. Schopf, J. W. (ed.) in *Earth's Earliest Biosphere: Its Origin and Evolution* (Princeton University Press, 1983).
3. Walker, J. C. G. & Brimblecombe, P. *Precamb. Res.* **28**, 205–222 (1985).
4. Froelich, P. N. et al. *Geochim. cosmochim. Acta* **43**, 1075–1090 (1979).
5. Cloud, P. *Econ. Geol.* **68**, 1135–1143 (1973).
6. Cameron, E. M. & Garrels, R. M. *Chem. Geol.* **28**, 181–197 (1980).
7. Gole, M. J. & Klein, C. J. *Geol.* **89**, 169–183 (1981).
8. Walker, J. C. G. et al. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 260–290 (Princeton University Press, 1983).
9. Walker, J. C. G. *Nature* **309**, 340–342 (1984).
10. Baur, M. E., Hayes, J. M., Studley, S. A. & Walter, M. R. *Econ. Geol.* **80**, 270–282 (1985).
11. Kasting, J. F., Zahnle, K. J. & Walker, J. C. G. *Precamb. Res.* **20**, 121–148 (1983).

Nuclear β -decay and the origin of biomolecular chirality

W. J. Meiring

Kernfysisch Versneller Instituut der Rijksuniversiteit,
Universiteitscomplex Paddepoel, 9747AA Groningen,
The Netherlands

The origin of homochirality in terrestrial life has long been debated¹. Asymmetric decomposition of racemic mixtures of biomolecules by longitudinally polarized β -rays has been described as a possible mechanism for inducing optical activity². Here I present an extension of former estimates given by Hegstrom³ with respect to this process. I derive the energy dependence and sign of the relative asymmetries in radiolysis cross-sections for both β^+ - and β^- -radiation, and the connection of these cross-sections with the decomposition rate constants. In the light of the amplification model devised by Kondepudi and Nelson^{4–7}, I compare the chiral selectivity of β -rays with that of weak neutral current effects in handed molecules^{8,9}. I argue that the exceptional prebiotic conditions required do not favour asymmetric β -radiolysis as the selector of the exclusive signature of optical activity in living nature.

Radiolysis can be regarded as the knocking-out of a bonding electron in a molecule with a typical bonding energy of a few eV. It has been shown³ that the asymmetry in this process of ionization can be described by means of a bound-helical-electron model, in which the asymmetry is a result of the helicity, h , of the electrons moving in molecular orbits. This helicity is predicted to exist in optically active molecules owing to the perturbation of the electron bound state by the spin-orbit interaction. It is of the order $h \sim \eta(\alpha Z)^2$, where Z is the atomic number of the chiral centre and η is a molecular asymmetry factor ($\sim 10^{-3}$ to 10^{-4} (ref. 10)), which takes into account the degree of asymmetry in the molecular structure.

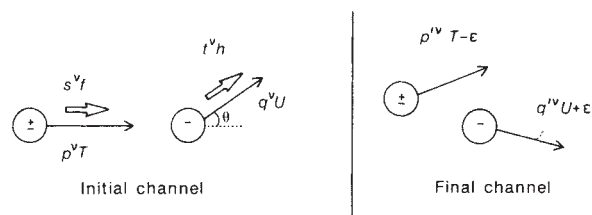


Fig. 1 Helicity-dependent free-lepton-electron scattering. The initial and final momentum four-vectors of the incoming particle (β^+ or β^-) are p^ν and p'^ν . Its spin four-vector is s^ν . The corresponding quantities of the molecular electron are q^ν , q'^ν and t^ν . The angle between $\vec{p}$ and $\vec{q}$ is θ . The particles have helicities f and h , and kinetic energies T and U . The energy ϵ , which will be transferred from the incoming to the molecular particle, can have the values $-U < \epsilon < T$. The angle ϕ (not shown) is the azimuth of $(\vec{p}' - \vec{q}')$ relative to $(\vec{p} + \vec{q})$.

For β^- -radiation, where the leptons are identical, it is the interference between the direct and exchange term that leads to a non-zero asymmetry³. The indistinguishability of the two electrons, causing the exchange effect, will be most important in collisions with large energy transfer. In this case the binding energy of the molecular electrons plays a secondary role and we can consider the electrons as free. This reduces the problem to the calculation of free-electron-electron (Møller) scattering. Kim and Inokuti¹¹ applied this free-electron approximation successfully to calculate corrections of ionization cross-sections for the exchange effect; here it will be used to obtain a helicity-dependent correction.

Figure 1 shows the scattering process. As the particles are longitudinally polarized, their spin four-vectors can be written as

$$s^\nu = \frac{f}{m} (p, (T+m)\hat{p}) \quad (1)$$

and similarly for t^ν . Here m is the electron rest mass, and for convenience I take $c = 1$. The total ionization cross-section is

$$\sigma^- = \int_I^{\frac{1}{2}(T-U)} d\epsilon \frac{d\sigma^-}{d\epsilon} \quad (2)$$

where I is the ionization energy of the molecular electron. The differential cross-section $d\sigma^-/d\epsilon$ can be written as the sum of a spin-independent and a spin-dependent term¹²:

$$\frac{d\sigma^-}{d\epsilon} = \frac{d\sigma^-}{d\epsilon} + fh \frac{d\sigma^-}{d\epsilon} \quad (3)$$

From (2) and (3) the helicity-dependent correction $\Delta\sigma^-$ in the ionization cross-section, averaged over all the directions of $\vec{q}$, is

$$\Delta\sigma^- = \frac{1}{2} \int_{-1}^1 d\cos\theta \int_I^{\frac{1}{2}(T-U)} d\epsilon \frac{d\sigma^-}{d\epsilon} \quad (4)$$

for the incident and molecular electrons having the same helicity ($fh = +1$), and $-\Delta\sigma^-$ for particles with opposite helicity ($fh = -1$).

For electrons the spin-dependent part of equation (3) is the sum of a direct term, an exchange term and an interference term^{13,14}:

$$fh \frac{d\sigma^-}{d\epsilon} = \frac{4m^2 r_0^2}{\sqrt{u(u-4m^2)}|\vec{p} + \vec{q}|} \int d\phi \left[\frac{A^-}{v^2} + \frac{B^-}{w^2} - \frac{C^-}{vw} \right] \quad (5)$$

with

$$A^- = -m^2 p' \cdot s q' \cdot t - m^2 v s \cdot t, \quad (6a)$$

$$B^- = -m^2 q' \cdot s p' \cdot t - m^2 w s \cdot t \quad (6b)$$

$$C^- = (w - m^2) p' \cdot s q' \cdot t + (v - m^2) q' \cdot s p' \cdot t + m^2 q \cdot s p \cdot t + \frac{1}{2}(vw - 2m^2(u - 2m^2))s \cdot t \quad (6c)$$

where $u = (p + q)^2$, $v = (p - p')^2$ and $w = (p - q')^2$ are the Man-

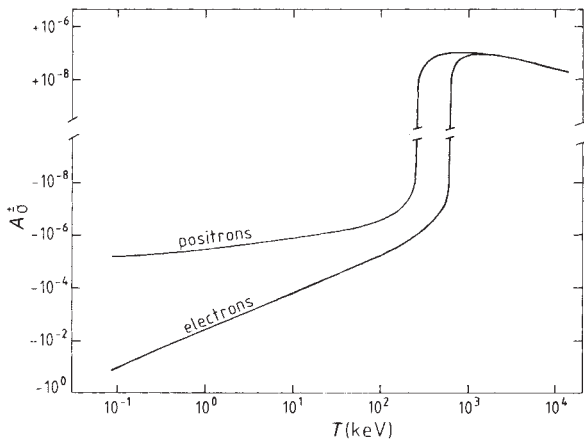


Fig. 2 Relative asymmetry $A_{\sigma}^{\pm}$ in ionization cross section plotted against kinetic energy T of the incident $\beta^{\pm}$ -particles for purely helical scattering and scattered particles. It appears that A_{σ}^{-} changes sign at ~ 590 keV and A_{σ}^{+} at ~ 240 keV. This change of sign is almost independent of the ionization energy I . At relativistic energies electrons and positrons behave identically. At lower energies the asymmetry A_{σ}^{-} of electrons dominates by far A_{σ}^{+} due to the Pauli exclusion principle, and approaches -1 for ionization energies.

delstam invariants with metric $(+, -, -, -)$ and $r_0 = e^2/m = 2.8$ fm. Now assume U and I to be much less than T and m , so they can be neglected. Then, by inserting (5) and (6) in (4) and evaluating (4) to first order in q gives:

$$\Delta\sigma^{-} = \frac{4\pi q r_0^2}{3p^3 T} \left[(T^3 + 2mT^2 - 2m^2T - 2m^3) \ln \frac{T}{I} - \frac{1}{4} T(T^2 - 2mT - 4m^2) \right] \quad (7)$$

The same method can be applied to ionization by positrons. The spin-dependent part of the Bhabha scattering cross-section, containing a direct term, a virtual annihilation term and a cross term, is given by (5) after the term in brackets is replaced¹⁴ by

$$\frac{A^{+}}{v^2} + \frac{B^{+}}{u^2} + \frac{C^{+}}{uv} \quad (8)$$

with

$$A^{+} = A^{-} \quad (9a)$$

$$B^{+} = (w - 2m^2)p' \cdot s q' \cdot t + (v - 2m^2)q' \cdot s p' \cdot t + m^2 q \cdot s p \cdot t + \frac{1}{2}(vw - 4m^4)s \cdot t \quad (9b)$$

$$C^{+} = (w - 3m^2)p' \cdot s q' \cdot t + (v - m^2)q' \cdot s p' \cdot t + m^2 q \cdot s p \cdot t + \frac{1}{2}(v + 2m^2)(w - 2m^2)s \cdot t \quad (9c)$$

Equation (2) has now to be integrated up to $\varepsilon = T$, because positrons and electrons are distinguishable. With the same approximations, (4) yields

$$\Delta\sigma^{+} = \frac{4\pi q r_0^2}{3p^3(T+2m)} \left[(T^3 + 4mT^2 + 2m^2T - 2m^3) \ln \frac{T}{I} - \frac{7T^4 + 36mT^3 + 44m^2T^2 - 12m^3T - 12m^4}{6(T+2m)} \right] \quad (10)$$

Figure 2 shows a plot of the relative asymmetries $A_{\sigma}^{\pm} = \Delta\sigma^{\pm}/\sigma^{\pm}$ in ionization cross-sections against the kinetic energy of the incident particle with $I = U = 13.6$ eV. For $\sigma^{\pm}$ the cross-sections for the ionization of the hydrogen atom in the non-relativistic Born approximation¹¹ have been used, extended at relativistic energies by integrating the Møller and Bhabha cross sections $d\sigma^{\pm}/d\varepsilon$ (ref. 15). The asymmetry A_{σ}^{-} of electrons agrees with that obtained in ref. 3 for energies ≤ 100 keV.

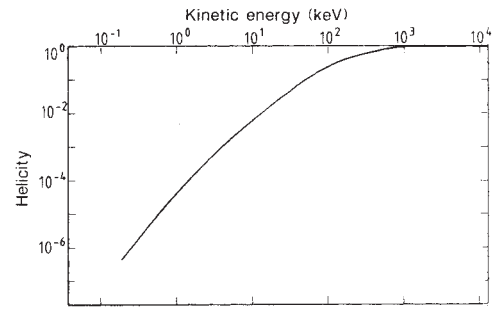


Fig. 3 Helicity-loss of electrons in water.

The radiolysis asymmetry is proportional not only to the molecular helicity, h , but also to the helicity, f , of the incoming leptons. For nuclear β -decay f has an initial value $\pm p/E$. During the slowing down of the particles in matter, this helicity is conserved in the relativistic energy region, where $\bar{s}$ follows $\bar{p}$. It vanishes, however, at non-relativistic energies, where $\bar{s}$ is frozen but $\bar{p}$ is randomized. We can define a function $H^{\pm}(T)$, being the helicity of an initial relativistic electron (positron) with initial helicity 1, when it has been slowed down to kinetic energy T . The function H obeys the differential equation¹⁶

$$\left\langle -\frac{dE}{ds} \right\rangle^{\pm} \frac{dH^{\pm}}{dT} = \frac{1}{2} \frac{d\langle\theta^2\rangle^{\pm}}{ds} \left(\frac{m}{T+m} \right)^2 H^{\pm} \quad (11)$$

Here $\langle -dE/ds \rangle$ is the average energy loss per unit path length (expressed in the Bethe-Bloch stopping-power formula¹⁵) and $d\langle\theta^2\rangle/ds$ is the mean square scattering angle per unit path length, which can be calculated from the Mott scattering cross-section¹⁵. Figure 3 shows the result of the numerical evaluation of (11) for water as the stopping medium.

What will be the joint influence of cross-section asymmetry, helicity loss and emitted β -ray spectrum to a final relative asymmetry $A_{\gamma}^{\pm} = \Delta\gamma^{\pm}/\gamma^{\pm}$ in the decomposition rate constant γ ? To answer this question I consider a primordial soup of water in which both a certain number of macromolecules and a β -ray source are dissolved. The strength of the source is S (in $s^{-1}m^{-3}$). The energy distribution of its emitted particles is $D(T)$ (eV^{-1}), for which we can take the normalized statistical factor in nuclear β -decay¹⁷

$$D(T) \sim pE(T - T_0)^2 \quad (0 < T < T_0) \quad (12)$$

T_0 being the end-point energy of the β -source. With N bonding electrons involved, γ (in s^{-1}) is given by

$$\gamma^{\pm} = NS \int_0^{T_0} dT \sigma^{\pm}(T) \left\langle -\frac{dE^{\pm}}{ds}(T) \right\rangle^{-1} \int_T^{T_0} dT' D(T') \quad (13)$$

and the asymmetry in γ by

$$A_{\gamma}^{\pm} = \pm \frac{hNS}{\gamma^{\pm}} \int_0^{T_0} dT \Delta\sigma^{\pm}(T) \times \left\langle -\frac{dE^{\pm}}{ds}(T) \right\rangle^{-1} H^{\pm}(T) \int_T^{T_0} dT' \frac{D(T')p'}{H^{\pm}(T')E'} \quad (14)$$

Figure 4 shows $A_{\gamma}^{\pm}$ as a function of the end-point energy T_0 of the β -source. I have assumed a value of 10^{-6} for the molecular helicity, h (taking $Z \approx 6$). The value of h may be reduced because the bonding electrons are shielded from the nucleus by the inner electrons. As (14) is actually a weighted average of all the bonding electrons, A_{γ} is expected to be smaller for large molecules with many non-helical bonds. In Fig. 4 a few relevant β -sources are also indicated. Note that β^{+} -sources with end-point energies smaller than 0.5 MeV scarcely exist, because electron capture will be the dominant process there. Most β^{+} -

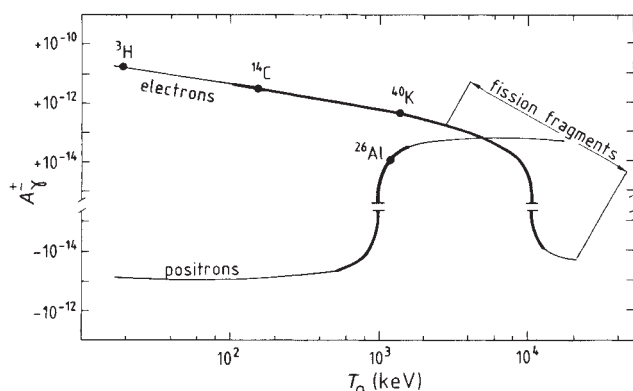


Fig. 4 Radiolysis asymmetry $A_{\gamma}^{\pm}$ plotted against end-point energy T_0 of the $\beta^{\pm}$ -sources with $h=+10^{-6}$. It turns out that for a β^{-} -source with $T_0 \sim 10$ MeV the contributions below and above 590 keV cancel, whereas A_{γ}^{+} for positrons changes sign at $T_0 \sim 1$ MeV. Furthermore, it appears that most β^{-} -sources, having an end-point energy $T_0 < 10$ MeV, cooperate with β^{+} -sources with $T_0 > 1$ MeV.

emitters are short-lived, and it is therefore unlikely that β^{+} -sources, other than possibly ^{26}Al , could have played a role in the selection of biomolecular chirality.

Kondepudi and Nelson⁴⁻⁷ have analysed an autocatalytic reaction mechanism that spontaneously breaks chiral symmetry, showing that the weak neutral current (WNC) effects are sufficient to determine the final outcome of optical activity. β -radiolysis can be incorporated into this model by taking the kinetic constant k_{-1} of the back reaction in the first step ($A + B \rightleftharpoons X_{L(D)}$) unequal, so that $\gamma A_{\gamma} = \frac{1}{2}(k_{-1L} - k_{-1D})$. The time development of the enantiomeric excess $\alpha = \frac{1}{2}(X_L - X_D)$ near the critical point (where the selection process takes place) is then given by (with omission of the statistical fluctuations)

$$\dot{\alpha} = Cg - \gamma A_{\gamma} X_c \quad (15)$$

(in M s^{-1}) where $C = (k_1 + k_2 X_c) A_c B_c + (k_{-1} + k_{-2} X_c) X_c$ and g is the WNC energy shift in units of kT ($\sim 10^{-17}$). The kinetic constants are the same as in ref. 7, and A_c , B_c and $X_c = \frac{1}{2}(X_L + X_D)_c$ are the concentrations of A, B and X at the critical point. To compete with the WNC effect, γ has therefore to be larger than $|Cg/A_{\gamma} X_c|$. With the numerical values of ref. 7 ($C = 5 \times 10^{-10} \text{ M s}^{-1}$, $X_c = 7 \times 10^{-4} \text{ M}$) this yields $\gamma > 7 \times 10^{-12} \text{ s}^{-1}$, which is too large in view of the strengths¹⁸ of the natural β -sources that could have been present in the prebiotic era. A decrease in C by a suitable change of the kinetic constants can obviate this, but will at the same time disfavour the selection by necessity. Hence this possibility can be ruled out.

It is also proposed¹⁹ that the reactants A and B are chiral as well, forming products of the corresponding chirality only in homochiral combinations. A relative asymmetry δ in the enantiomeric concentrations of A and B, caused by asymmetric β -radiolysis, will be equivalent to an inequality in rate constants. This asymmetry δ is kept constant by an influx of A and B from a large reservoir, where the radiolysis occurred before the autocatalytic process. Its asymptotic value, $\delta = -\gamma A_{\gamma}/(\kappa + 2\nu)$, will have been approached after a time $t > (\kappa + 2\nu)^{-1}$ (ref. 20), where κ is the total decomposition constant and ν the racemization constant. For the β -sources in the ocean ($\gamma = 5 \times 10^{-19} \text{ s}^{-1}$ (ref. 18)) and for alanine ($\nu = 8 \times 10^{-15} \text{ s}^{-1}$ at 0°C (ref. 20)) asymmetric β -radiolysis will only compete with WNC effects for irradiation times > 4 Myr. For shorter times, for $\kappa > \nu$ (for example when A and B are formed in a reversible reaction) and at higher temperatures, δ will be smaller. In addition, large amounts of A and B (10^7 mol or more⁷) need to be irradiated, both to overcome the statistical fluctuations in δ and to provide a reservoir large enough to control and stabilize the concentrations of A and B in the autocatalytic system. For this reason

stronger but more local β -sources are inadequate. The total crustal average¹⁸ is about 20 times the oceanic activity. This factor will, however, be lost because β -radiolysis in this case is a surface effect instead of a volume effect. A similar argument holds for uranium detrital deposits in streams and along beaches¹⁸, where the β -source is localized in a relatively small region. Fossil fission reactors fuelled by natural uranium¹⁸ are even more concentrated and occur rarely, and there is so far no reason to believe that life could have started exclusively in such an extreme environment. The irradiation of amino acids formed in the presolar nebula by ^{26}Al (ref. 18), accumulated by the decay of novae or supernovae, remains as a possible but speculative time model.

Based on these considerations, the interesting Vester-Ulbricht hypothesis cannot be rejected conclusively. However, as it demands very specific additional conditions for its validity this process does not seem to be a very appealing explanation for the choice of optical activity in living nature, compared with the more universal and omnipresent weak neutral current effects.

I thank J. v. Klinken, H. A. Tolhoek and D. Atkinson for helpful discussions. This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie with financial support from the Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek.

Received 30 April; accepted 20 August 1987.

1. Mason, S. F. *Nature* **311**, 19-23 (1984).
2. Ulbricht, T. L. V. & Vester, F. *Tetrahedron* **18**, 629-637 (1962).
3. Hegstrom, R. A. *Nature* **297**, 643-647 (1982).
4. Kondepudi, D. K. & Nelson, G. W. *Phys. Rev. Lett.* **50**, 1023-1026 (1983).
5. Kondepudi, D. K. & Nelson, G. W. *Physica A* **125**, 465-496 (1984).
6. Kondepudi, D. K. & Nelson, G. W. *Phys. Lett. A* **106**, 203-206 (1984).
7. Kondepudi, D. K. & Nelson, G. W. *Nature* **314**, 438-441 (1985).
8. Mason, S. F. & Tranter, G. E. *J. chem. Soc. chem. Commun.* 117-119 (1983).
9. Tranter, G. E. *J. chem. Soc. chem. Commun.* 60-61 (1986).
10. Hegstrom, R. A., Rein, D. W. & Sanders, P. G. H. *J. chem. Phys.* **73**, 2329-2341 (1980).
11. Kim, Y. K. & Inokuti, M. *Phys. Rev. A* **3**, 665-678 (1971).
12. Bincer, A. M. *Phys. Rev.* **107**, 1434-1438 (1957).
13. Raczka, R. & Raczka, A. *Bull. Acad. polon. Sci.* **6**, 463-468 (1958).
14. Holzwarth, G. Z. *Phys.* **191**, 354-359 (1966).
15. Rohrlach, F. & Carlson, B. C. *Phys. Rev.* **93**, 38-44 (1954).
16. Bouchiat, C. & Levy-Leblond, J. M. *Nuovo Cim.* **33**, 193-200 (1964).
17. Cohen, B. L. *Concepts of Nuclear Physics* (McGraw-Hill, New York, 1971).
18. Hegstrom, R. A., Rich, A. & van House, J. *Nature* **313**, 391-392 (1985).
19. Hegstrom, R. A. *Nature* **315**, 749-750 (1985).
20. Hegstrom, R. A. *Origins of Life* **14**, 405-411 (1984).

Migration rates of human populations from surname distributions

A. Piazza*, S. Rendine*, G. Zei†, A. Moroni‡ & L. L. Cavalli-Sforza§

* Dipartimento di Genetica, Biologia e Chimica Medica, Università di Torino, Via Santena 19, Torino 10126, Italy
† Laboratorio di Genetica e Biochimica Evoluzionistica, CNR, Pavia, Italy

‡ Istituto di Ecologia, Università di Parma, Parma, Italy

§ Genetics Department, Stanford University, Stanford, California 94305, USA

Migration is an important factor in the biological evolution of human populations, and surnames provide one of the simplest records of identification. The distribution of surnames can supply quantitative information on the structure of human populations¹⁻³. Surnames considered as alleles of a gene transmitted only by the male line can be assumed to be neutral markers⁴ and therefore satisfy the expectations of the neutral theory of evolution⁵, which is entirely described by random genetic drift, mutation and migration. As data on surnames are easier to collect than those from genes, the information yield is potentially increased, but the validity of the conclusions must be tested in actual samples. The purpose

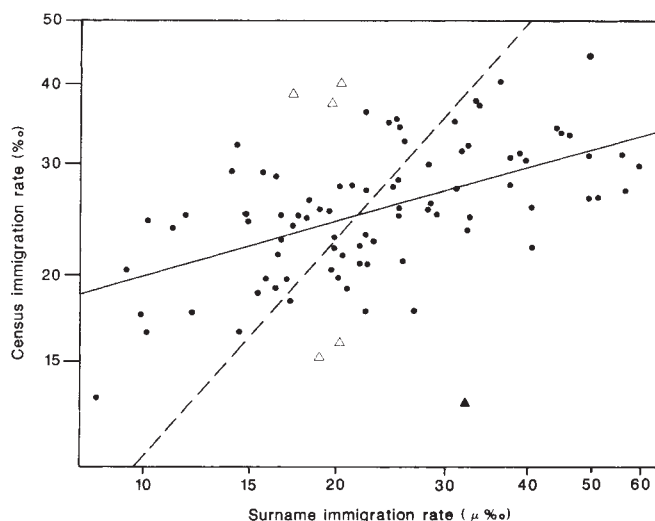


Fig. 1 Immigration rates of 91 Italian provinces estimated from telephone directories using surname data, as explained in the text, are plotted against immigration rates from Italian official demographic data (years 1967–1980). Both axes are represented on a logarithmic scale. A regression analysis was performed where the surname immigration rate was defined as the predictor variable (abscissa, x) and the census immigration rate as the response variable (ordinate, y). The general model $y = ax^b$, passing through the origin where the prediction is expected to be zero, was tested. The logarithmic formulation of the model $\log y = \log a + b \log x$ makes it possible to estimate a and b and test their statistical difference from unity; if $b = 1$, the immigration rate from surnames is proportional to the census immigration rate, and if a also equals 1, the two are the same. The results of the regression analysis are shown. For $a = 1$ (dotted line), the estimate of b (0.963 ± 0.011) is close to 1 (linearity) and the fit to the regression is highly significant ($F_{1,89} = 8091$, $r^2 = 0.989$). If a is also estimated (solid line), we find $b = 0.269 \pm 0.049$, $\log a = -1.159 \pm 0.081$, and the fit is less satisfactory ($F_{2,88} = 30.19$, $r^2 = 0.355$), but this surprising result probably depends on the range over which the fit is computed. Open triangles: outlier data points, as defined by the analysis of the studentized residuals¹¹. The most significant of them (Trieste, filled triangle) has been excluded from the regression analysis.

of this report is to compare the estimates of migration rates in Italy, as inferred by the surname distribution found in the telephone directories and other sources, with the corresponding estimates from official demographic sources. Our findings show that in these samples the ratio of surnames to individuals makes it possible to calculate reliable estimates of migration rates.

As most surname can be considered as neutral alleles of a genetic locus, they can be analysed using the theory of Karlin and McGregor⁶, which gives the distribution of neutral alleles expected in a population of N haploid individuals, each carrying one of k different alleles (surnames, in our case), subject to a process of death at random. Dead individuals are replaced by new ones carrying the same allele (the same surname) or externally from mutation and immigration, occurring with rate ν . This model has been shown to fit the observed surname distribution of the Parma valley in Italy². Because ν represents the sum of the surname mutation rate (usually quite rare) plus the immigration rate (several orders of magnitude higher), it is reasonable to assume ν represents the simple immigration rate. Moreover the Italian population, like any other population in developed countries, is relatively stable over a short period of time, so that the amount of immigration is roughly equivalent to the amount of emigration. Therefore the rate ν provides an estimate of the migration rate.

It has been shown⁴ that the distribution, originally given by Fisher⁷, for estimating the species abundance from a representative sample of fauna is a limit of the Karlin-McGregor distribution when the number of organisms (in our case, individuals) is much higher than the number of species (in our case, surnames). As the last condition holds for most practical cases and the parameters of the Fisher distribution have interesting properties⁸, we will adopt this procedure.

If N is the total number of individuals who form the sample and S is the number of surnames found in that sample, the following system of non-linear equations

$$N = \alpha(1 - \nu)/\nu \quad (1a)$$

$$S = -\alpha \log \nu \quad (1b)$$

allows ν , the immigration rate, and the parameter α to be obtained, given N and S . The quantity α is a useful constant of the model, and measures the 'density' of a surname in a given sample of individuals which is not a function of its size, N (ref. 7); values of α from different samples of surnames collected in the same population or society are expected to be equal. On this assumption, α can be estimated by equations (1) from any collection of surnames, for instance our Italian telephone directories; then equation (1a) can be used to estimate the immigra-

tion rate, ν , by substituting the parameter α and the sample size of individuals, N , referred to for the rate ν (in our application this is the census size which differs from the number of telephone users).

Our results are obtained by applying this approach to two large sources of data, and by correlating the immigrations estimated from the study of surnames with values from standard demographic sources. A primary source of data was the telephone directory from 91 Italian provinces, commercial subscribers having been eliminated.

Because in the model formulation, α does not depend on the sample size of the telephone users, but on the abundance of surnames, immigration was estimated from (1a), rewritten as

$$\nu = \alpha / (P_m + \alpha) \quad (2)$$

The parameter α was taken from the solution of system (1), with S and N being the number of surnames and the number of telephone users provided for each province of Italy by the telephone directories. P_m is the male census size from the same provinces; this was obtained by halving the number of people officially registered in each province of Italy at the end of 1978. The immigration rates ν , as estimated from (2), have been compared with actual immigration rates, calculated by averaging the yearly ratios 1967–1980 of newly registered individuals to the total residents in the 91 Italian provinces, as published by the Italian registration offices⁹. The hypothesis that these two sets of immigration rate values give equivalent results has been tested.

The surnames collected from the Italian telephone directories and forming the basis of this work have been summarized by De Felice¹⁰, who presents tabulations by different geographic subdivisions of the most frequent surnames. All Italian telephone directories are published yearly by one centralized company, SEAT. We had access to their computerized files which included the telephone users for 1978, the year considered by De Felice.

Figure 1 shows the correlation between observed and estimated migration coefficients. The Pearson correlation coefficient is 0.596 ± 0.083 , significantly different from zero. The influence curve and the studentized residuals¹¹ have been analyzed to detect possible outliers. Of these putative outliers displayed in the Fig. 1, the most extreme (Trieste) was only recently annexed to Italy; history and tourism may be influential for Venice and Genoa (which deviate in the same direction as Trieste) and could be responsible for an anomalous pattern of migration. The interpretation of the other outliers corresponding to the towns of Nuoro, Caserta, Avellino (all above the line in Fig. 1) is still to be investigated; all three belong to regions (Sardinia

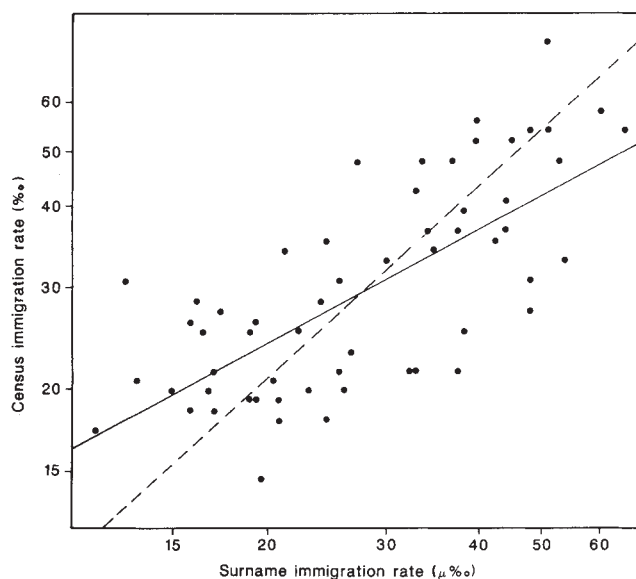


Fig. 2 Plot of immigration rates of 59 Italian geographic areas estimated from 1936–1964 consanguinity dispensations, using surname data as explained in the text, against immigration rate from Italian demographic data (year 1961). Both axes are represented on a logarithmic scale. The regression analysis has been performed as described in the legend to Fig. 1. For $a=1$ (dotted line), the estimate of b (1.009 ± 0.013) does not differ significantly from 1 (linearity), and the fit to the regression is highly significant ($F_{1,58} = 6235$, $r^2 = 0.991$). If a is also estimated (solid line), then $b = 0.626 \pm 0.080$, $\log a = 0.563 \pm 0.117$; the fit is less satisfactory ($F_{2,57} = 60.56$, $r^2 = 0.520$) for the same reasons as in Fig. 1.

and Campania) where emigration was notably higher than immigration during the period considered.

As we wish to predict the immigration rate from ν , a regression analysis was performed where the surname immigration rate was defined as the predictor variable (x) and the census immigration rate as the response variable (y). The general model $y = ax^b$, which passes through the origin (where the prediction is expected to be zero), was tested and the results of the analysis are described in the legend to Fig. 1. The linear regression fit is statistically highly significant, but a nonlinear regression, even though less satisfactory, is also compatible with the data.

A relevant point to be considered is whether the telephone directories contain a non-random sample of the population. Our Italian sample has 10,473,727 registered telephone users with 59,961,000 inhabitants, or one telephone for every 5.44 residents¹⁰; the average family size is about 4 and therefore most families are represented. Where the telephone directories are known to contain a non-random sample of the population (for instance individuals from the highest social classes), the same technique could be applied using other collections of listed surnames, such as those held by electricity-supplying companies, because today most people use electricity, irrespective of their social or economic class.

The second body of data originates from a list of about 540,000 consanguineous marriages for which dispensations were requested in Italy between the years 1910 and 1964 (ref. 4). Only the surnames of brides were used. Consanguinity dispensations are collected by dioceses (280 in number), which tend to be larger than provinces in northern Italy, and much smaller in the central and southern regions. Dioceses and provinces frequently have non-overlapping boundaries. By agglomerating neighbouring dioceses into geographic areas chosen to correspond as closely as possible to provinces or province clusters, 59 geographic areas were defined. The migration data is again derived from the number of new residents per year per province, expressed as percentages of the resident population. The consanguinity data refer to a population which tends to be more rural than that of the telephone data, and the relevant migration refers to an earlier and less-defined period. Probably there is also a selection for a more conservative, less mobile population.

Figure 2 shows the immigration values estimated from surnames of consanguineous dispensations for years 1936–1964 versus those from official data for the year 1961. The correlation is $r = 0.721 \pm 0.134$ on 59 geographic areas. An earlier date for migration has been also tested because consanguineous mar-

riages are selected on the basis of migration of parents or more remote ancestors¹². Using official immigration data for 1911, the correlation is almost the same ($r = 0.713 \pm 0.137$), but using the same 1911 official immigration data and dispensations for years 1911–1936, the correlation is smaller ($r = 0.398 \pm 0.162$). Correlation coefficients can be strongly affected by outliers but none of these were eliminated from our calculations. With consanguinity dispensations, the estimate of ν is obtained by employing the sample size as P , rather than the census size. This seems to be a better procedure when there are substantial deviations from proportionality between the two. The results of the regression analysis are similar to the previous ones from the telephone data and are shown in Fig. 2.

Our findings show that, provided a sample of adequate size is available, the ratio of surnames to individuals in the sample makes it possible to calculate an estimate of male immigration rates. The density of particular Italian surnames does not place any limitation on the application of the technique. The model is independent of the number of surnames but its accuracy depends on the number of individuals carrying the surnames, whatever the number of surnames might be. The number of individuals is very high when compared to the usual size of test samples, taken for blood groups for example, and this is one of the potential advantages of the technique. Further investigation is needed to refine these estimates, which are simple to collect and could be very useful for studies of population structure.

We are most grateful to SEAT for allowing access to their telephone-user files. Financial support was from the US National Institutes of Health and MPI, Italy. Computer time has been provided by CSI-Piemonte, Italy.

Received 8 July; accepted 8 September 1987.

1. Crow, J. F. & Mange, A. P. *Eugen Quart.* **12**, 199–203 (1965).
2. Yasuda, N., Cavalli-Sforza, L. L., Skolnick, M. & Moroni, A. *Theor. Populat. Biol.* **5**, 123–142 (1974).
3. Lasker, G. W. in *Surnames and Genetics Structure* (Cambridge University Press, 1985).
4. Zei, G., Guglielmino-Matessi, R., Siri, E., Moroni, A. & Cavalli-Sforza, L. L. *Ann. hum. Genet.* **47**, 329–352 (1983).
5. Kimura, M. in *The Neutral Theory of Evolution* (Cambridge University Press, 1983).
6. Karlin, S. & McGregor, J. *Proc. 5th Berkeley Symp. Math. Stat. Prob.* **4**, 415–438 (1967).
7. Fisher, R. A. *J. Anim. Ecol.* **12**, 42–58 (1943).
8. Zei, G., Piazza, A., Moroni, A. & Cavalli-Sforza, L. L. *Ann. hum. Genet.* **50**, 169–180 (1986).
9. ISTAT *Popolazione e movimento anagrafico dei Comuni* (Istituto Poligrafico dello Stato, Rome, 1967–1970, 1972–1974, 1976, 1978, 1980).
10. De Felice, E. in *I cognomi italiani* (Mulino, Bologna, 1978).
11. Cook, R. D. & Weisberg, S. in *Residuals and Influence in Regression* (Chapman & Hall, New York, 1982).
12. Cavalli-Sforza, L. L., Kimura, M. & Barrai, I. *Genetics* **54**, 37–60 (1966).

Bill size polymorphism and intraspecific niche utilization in an African finch

Thomas Bates Smith

Museum of Vertebrate Zoology, University of California, Berkeley, California 94720, USA

The African estrildid finch, *Pyrenestes ostrinus*, shows a non-sex-linked polymorphism of bill size. Within *P. ostrinus* and its congeners, *P. sanguineus* and *P. minor*, variation in bill size is extreme, bimodally distributed and not determined by sex, age, body size or geographical origin. *P. ostrinus* appears to mate randomly with respect to bill size. Bill morphs differ with respect to diet and feeding performance on hard and soft seeds of sedges, *Scleria* spp. (Cyperaceae), the bird's principal food. The novel pattern of intraspecific divergence in trophic structure contrasts sharply with patterns of morphological variation in Darwin's finches (*Geospiza*) and other avian species not showing bill size polymorphisms. Coefficients of variation (CVs) for bill size are extraordinarily large while CVs for body size are small and typical of other passerine species. Differences in bill size between morphs exceed those reported for several sympatric species of Darwin's finches believed to show character displacement¹, and like many co-occurring congeneric species the morphs differ in diet. Discrete intrapopulation variation in the trophic structures of birds usually consists either of differences between the sexes (sexual dimorphism)², or of morphological asymmetries, such as bill crossing^{3,4}. Here I describe a third pattern of discrete variation previously hypothesized to occur in the hook-billed kite (*Chondrohierax uncinatus*)⁵, and found only in a few species of other vertebrates⁶⁻¹⁰, in which size differences in trophic apparatus are not determined mainly by sex. Polymorphisms of trophic apparatus have been correlated with differential resource utilization and, despite the scarcity of reports, have been postulated to be important in reducing intraspecific competition and in promoting speciation⁶.

Historically, the bimodally distributed bill size of *Pyrenestes* has perplexed taxonomists. All three species, restricted to the tropical forests of Africa, are reported to exhibit distinct large- and small-billed variants within sex (Fig. 1) and age classes. This pattern shows no clear correlation with geographical origin¹¹⁻¹³. Evidence that large- and small-billed forms can interbreed has also been reported¹³.

I conducted two years of field research (September–December 1983, January–May 1985, August 1985–November 1986) on the black-bellied seed-cracker (*Pyrenestes ostrinus*) in south-central Cameroon. Over 1,400 individuals were netted, banded and measured and their breeding biology examined. Each bird was banded with a unique, randomly chosen combination of a numbered aluminium band and three coloured bands to allow individual identification in the field. Measurements on 10 characters were taken for each bird (Table 1). To gain additional information on variation in the three species, these same characters were measured on 840 museum specimens.

Analysis of mensural characters for individuals trapped in the field showed significant differences between males and females (Table 1). Females tend to be smaller in overall body size than males. Juveniles up to 4 months of age tend to be smaller than adults. All bill measurements were bimodally distributed or their distributions were significantly skewed. In particular, lower mandible width (LMW) was bimodally distributed and actually discontinuous within most sex (Fig. 1) and age classes. Distributions of wing length were not significantly different from normal (Kolmogorov–Smirnov goodness of fit $P > 0.1$). CVs for wing length were 3% in both sexes, whereas values for the six bill characters were nearly four times larger

Table 1 Coefficients of variation for mensural characters of adult *Pyrenestes ostrinus* captured on the study area

Character	Males			Females		
	No.	Mean (g or mm)	CV (%)	No.	Mean (g or mm)	CV (%)
Mass	720	19.3	8	463	18.3	8
Wing	710	61.2	3	451	60.2	3
Tail	675	51.9	5	431	50.9	5
Tarsus	718	16.5	5	456	16.3	5
Lower mandible width	724	13.6	10	462	13.2	11
Upper mandible width	716	8.7	11	457	8.4	11
Bill depth	696	11.0	12	444	10.6	12
Upper mandible length	720	10.0	8	460	9.9	9
Lower mandible length	717	8.8	9	456	8.6	9
Upper mandible depth	717	4.0	15	456	3.8	16

Male and female *P. ostrinus* were sexed on the basis of plumage¹¹. Males were significantly larger than females in all characters on the basis of Kolmogorov–Smirnov two sample tests, $P < 0.001$. All character distributions were significantly different from normal in both sexes except wing in males, wing and tarsus in females (Kolmogorov goodness of fit tests, $P < 0.001$). Within-sex indices of body size (wing cord and cube root of body mass) were significantly correlated with measures of bill size (Spearman rank correlation, $P < 0.001$). Measurements were taken using dial calipers except for mass which was measured using a 50-g Pesola spring scale. Measurements were taken as follows: wing length, from carpal joint to the tip of the longest primary; tail length, from uropygial gland to tip of longest rectrix; tarsus length, from tibiotarsal joint to distal undivided scute; lower mandible width measured at base of bill; upper mandible width at nostril; bill depth in vertical plane level at the anterior edge of the nares; upper mandible length, from anterior edge of nostril to tip; lower mandible length, from proximal midpoint to the tip; upper mandible depth, from ventral edge of nostril to edge of upper mandible.

(Table 1). Analysis of museum specimens within defined passerine avifaunal zones¹⁴ showed a similar pattern for all species. CVs for bill size are larger than those for six species of Darwin's finches previously believed to show the largest CVs amongst birds¹⁵. In most passerine species, CVs for bill and body size range from 3–5% (ref. 16). In *Pyrenestes*, bill-size variation is considerably greater than this, but body-size variation is not. A similar pattern of variation in the snail-eating hook-billed kite led workers to hypothesize that it was polymorphic for bill size, but the possibility that more than one species of kite was involved could not be ruled out⁵.

There is no evidence for geographical segregation or local habitat segregation between large- and small-billed birds in any of the three species. Large- and small-billed forms occur together in all populations of the three presently recognized species in Africa (Fig. 1).

At the field site in Cameroon, *P. ostrinus* shows two peaks of breeding that coincide with major and minor rainy seasons. During the 1984 and 1985 breeding seasons, approximately 15% of 60 nests fledged offspring. Both parents were measured at 19 nests. They appear to have mated randomly with respect to bill size (Fisher's exact test $P = 0.22$). At 19 additional nests the bill size of one parent was measured and that of the other was estimated using binoculars or a spotting scope. For both samples combined, pairings remained random (Fisher's exact test $P = 0.12$). Observed pairings, when compared with expected pairings based on proportions of adult morphs on the study area during the breeding season estimated from mist netting, also show random mating with respect to bill size ($\chi^2 = 0.36$, $P > 0.54$; $\chi^2 = 0.79$, $P > 0.67$, for each sample, respectively). Furthermore, there is no significant sexual bias in the bill size of mixed pairs, that is, males are not always larger than females.

Indirect evidence suggests the bill dimorphism has a genetic

basis. First, the distribution of LMW in juveniles less than 3 weeks after fledging is bimodal. If differential bill wear were a factor contributing to this dimorphism, the discontinuity would not be expected to show up at such an early age. Birds do not appear to engage in any behaviour that would impart wear to the basal portion of the lower mandible. Furthermore, of 51 birds measured and subsequently retrapped and measured again as much as three years later, no significant change in bill size was detected, although bills of juveniles grow slightly for several months after fledging. Although it has been difficult to examine broods at fledging in the field, breeding experiments now underway in captivity at the Riverbanks Zoological Park, Columbia, South Carolina, suggest that mixed pairs may produce offspring with both large and small bills but not bills of intermediate size. For example, in the only successful breeding of *P. ostrinus* in captivity, a large-billed male (LMW=15 mm) mated with a small-billed female (LMW=12.6 mm). The pair fledged three offspring: a large-billed bird (LMW=15.2 mm) and two small-billed birds (LMW=12.6 mm and 11.9 mm). Young were measured approximately two months after fledging, when their bills had reached nearly adult size.

There is an ecological correlation between sedge consumption, feeding performance, and bill size (Table 2). *P. ostrinus* specializes on the seeds of sedges (*Scleria* spp.). Large-billed morphs readily crack large, hard seeds while small-billed morphs do so with difficulty and prefer not to feed on them. Small-billed morphs have higher feeding efficiencies on small, soft seeds than do large-billed morphs. These differences in behaviour are consistent with dietary differences observed in the field (Table 2). Dietary differences between morphs are similar to those found between congeneric, sympatric species of Darwin's finches¹⁶ and appear to be one of the few examples among vertebrates where a polymorphism is associated with differential niche utilization.

This novel pattern of divergence in trophic apparatus raises important evolutionary questions. Firstly, why did trophic divergence occur by means of an intrasexual polymorphism, rather than through the more frequent mechanism of enhanced sexual dimorphism? Secondly, what factors allowed populations of *Pyrenestes* to diverge polymorphically across a continent already exhibiting a high diversity of granivorous birds¹⁷, and with a presumably high potential for interspecific competition? The difference in mean bill size between morphs, expressed as a percentage of the mean of the small morph, is 23%, which is larger than the 15% minimum described for sympatric species of Darwin's finches and other avian species believed to show character displacement^{1,18,19}. The correlations among feeding performance, observed food preferences and bill size imply that

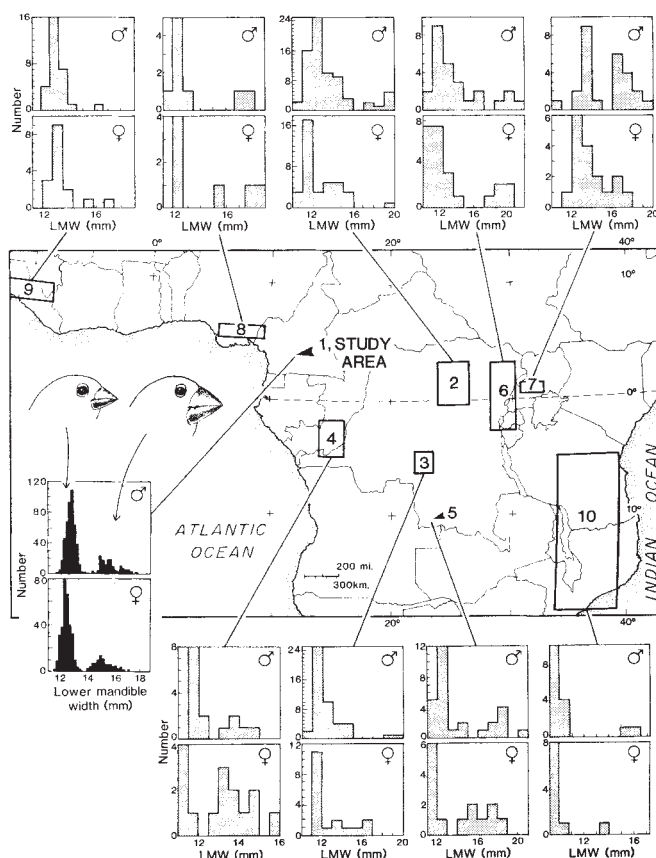


Fig. 1 Distributions of lower mandible width for adult males and females within sampling areas 1-10. 1, *P. ostrinus* captured on study area in Cameroon; 2-8, museum specimens; 9, *P. sanguineus*; and 10, *P. minor* based on museum specimens. All measurements were made by the author except for two specimens of *P. minor* reported elsewhere¹⁰.

trophic factors may play a major part in maintaining the polymorphism. Large seasonal differences in rainfall or a pattern of sporadic droughts could affect the abundance of certain sedge seeds resulting in differential survival of morphs, especially if their abilities to switch to other foods are limited. Similar correlations with the occurrence of drought resulting in intense natural selection on bill size in populations of *Geospiza* have recently been reported^{1,16,20}.

Table 2 Sedge use and feeding performance of morphs

	Sedge A				Sedge B			
	N_1	N_2	Feeding time (s) $\bar{X} \pm \text{s.e.m.}$	% Crops	N_1	N_2	Feeding time (s) $\bar{X} \pm \text{s.e.m.}$	% Crops
Small-billed morph	11	2	25 $\pm$ 2.5	29	23	23	6 $\pm$ 0.25	36
Large-billed morph	11	11	18 $\pm$ 2.5	83	25	25	9 $\pm$ 0.32	13

Sedge A is the hard-seeded sedge *Scleria verrucosa*. Seed subsamples of sedge A used in feeding trials were taken from a sample with mean hardness and diameter of 153 Newtons and 3.2 mm, respectively. Sedge B is the soft-seeded sedge *Scleria* sp. (aff. *achtenii*, D. Thomas, personal communication), mean hardness and diameter of random samples collected on the study area were 13 Newtons and 2.9 mm, respectively. 7 g of sedge A was used in each trial and 10-15 stalks of sedge B. Each wild-caught adult bird was placed in a 1 $\times$ 1 $\times$ 0.5 m cage for testing. N_1 , number of individuals which could be sustained on a commercial seed diet (birds which refused to feed in captivity were released). N_2 , number of individuals feeding on sedge. The time it took for a bird to feed on an individual seed was recorded to the nearest second using a stop watch. Average time to feed was calculated from 5 to 101 observations of an individual bird cracking a seed, then averaged across individuals. Small morphs preferred not to feed on sedge A while all large morphs fed on sedge A, Fisher's exact test, $P = 0.0001$. Large morphs had a higher efficiency than small morphs on sedge A but the difference was not significant, probably due to small sample size. Both morphs fed readily on sedge B, but small morphs fed significantly more efficiently than large morphs, $t = -6.63$, $P < 0.001$. Diets between morphs were significantly different $\chi^2 = 68.83$, d.f. = 1, $P < 0.0001$. Per cent crop, number a given morph netted with sedge A or B in crop/total number of same morph that had food in crop. One record was used for each individual and includes 608 individuals netted from November 1985 through October 1986, sex and age classes combined. Individuals with both sedge A and B in their crops constituted less than 3% of the samples and were ignored.

I thank the Government of the Republic of Cameroon for permission to conduct the field research; the curators and staff at the following museums: Koninklijk Museum voor Midden-Afrika; British Museum; American Museum of Natural History; Museum für Naturkunde der Humboldt-Universität zu Berlin; Zoologisches Forschungsinstitut und Museum Alexander Koenig; Field Museum; Musée National d'Histoire Naturelle; L.A. County Museum. I thank P. Grant, R. Johnston, A. Meyer, J. Patton, F. Pitelka, G. Roderick, R. Sage, D. Schluter, D. Wake and P. Williams and two anonymous reviewers for critical comments on the manuscript; F. Pitelka, P. Smith for advice and technical assistance; my field assistants; and the Riverbanks Zoological Park, Columbia, SC. The research was supported by grants from the National Geographic Society, NSF, American Ornithologists' Union, National Academy of Science, and Museum of Vertebrate Zoology, University of California, Berkeley.

Received 8 June; accepted 18 August 1987.

1. Schluter, D., Price, T. D. & Grant, P. R. *Science* **227**, 1056-1059 (1985).
2. Selander, R. K. *Condor* **68**, 113-151 (1966).
3. Richards, L. P. & Bock, W. L. *AOU Orn. Monogr.* **15** (1973).
4. Knox, A. G. *Bull. Br. Orn. Club.* **103**, 114-118 (1983).
5. Smith, T. B. & Temple, S. A. *Auk* **99**, 197-207 (1982).
6. Liem, K. F. & Kaufman, L. S. in *Evolution of Species Flocks* (eds Echelle, A. A. & Kornfield, I.) (University Maine Press, 1984).
7. Collins, J. P. & Holomuzki, J. R. *Can. J. Zool.* **62**, 168-174 (1984).
8. Sage, R. D. & Selander, R. K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4669-4673 (1975).
9. Kornfield, I., Smith, D. C., Gagnon, P. S. & Taylor, J. N. *Evolution* **36**, 658-664 (1982).
10. Orton, G. R. *Copeia* **97-100** (1954).
11. Chapin, J. P. *Bull. Am. Mus. Nat. Hist.* **49**, 415-441 (1924).
12. Vincent, J. *Ibis* **6**, 98-101 (1936).
13. Chapin, J. P. *Bull. Am. Mus. Nat. Hist.* **75B**, 489-498 (1954).
14. Crowe, T. M. & Crowe, A. A. *Zool. Lond.* **198**, 417-442 (1982).
15. Grant, P. R., Abbott, L., Schluter, D., Curry, R. L. & Abbott, L. K. *Biol. J. Linn. Soc.* **25**, 1-39 (1985).
16. Grant, P. R. in *Ecology and Evolution of Darwin's Finches* (Princeton University Press, 1986).
17. Keast, A. in *Proc. 16th Int. Orn. Congr.* (Brill, Leiden, 1972).
18. Grant, P. R. *Syst. Zool.* **17**, 319-333 (1968).
19. Grant, P. R. *Syst. Zool.* **21**, 289-291 (1972).
20. Boag, P. T. & Grant, P. R. *Science* **214**, 82-85 (1981).

Oscillations of cytosolic Ca^{2+} in pituitary cells due to action potentials

Werner Schlegel*, Benoît P. Winiger*, Patrice Mollard†§, Pierre Vacher‡, François Wuarin*, Gaston R. Zahnd*, Claes B. Wollheim† & Bernard Dufy‡

* Fondation pour Recherches Médicales,

† Institut de Biochimie Clinique, Department of Medicine, University of Geneva, CH-1211 Geneva 4, Switzerland

‡ Laboratoire de Neurophysiologie, UA CNRS 1200, et § UFR de Sciences Pharmaceutiques, Université de Bordeaux II, Bordeaux, France

Electrical activity in non-neuronal cells can be induced by altering the membrane potential and eliciting action potentials. For example, hormones, nutrients and neurotransmitters act on excitable endocrine cells¹. In an attempt to correlate such electrical activity with regulation of cell activation, we report here direct measurements of cytosolic free Ca^{2+} changes coincident with action potentials. This was achieved by the powerful and novel combination of two complex techniques, the patch clamp² and microfluorimetry using fura 2 methodology^{3,4}. Changes in intracellular calcium concentration were monitored in single cells of the pituitary line GH₃B₆. We show that a single action potential leads to a marked transient increase in cytosolic free calcium. The size of these short-lived maxima is sufficient to evoke secretory activity. The striking kinetic features of these transients enabled us to identify oscillations in intracellular calcium concentration in unperturbed cells resulting from spontaneous action potentials, and hence provide an explanation for basal secretory activity. Somatostatin, an inhibitor of pituitary function⁵, abolishes the spontaneous spiking of free cytosolic Ca^{2+} which may explain its inhibitory effect on basal prolactin secretion. Our data therefore demonstrate that electrical activity can stimulate Ca^{2+} -dependent functions in excitable non-neuronal cells.

Following the firing of current-induced action potentials, marked changes in intracellular calcium concentration ($[\text{Ca}^{2+}]_i$), apparent from the fura 2 fluorescence ratio (F_{350}/F_{380}), were observed (Fig. 1a, b). $[\text{Ca}^{2+}]_i$ increased concomitantly with the action potential to a maximum reached within 200-400 ms and was reduced to basal levels within a few seconds. Magnitude and duration of the $[\text{Ca}^{2+}]_i$ transients depended on the number of action potentials fired, and were independent of the duration of the depolarizing pulse; firing of several action potentials in a short interval led to larger alterations of $[\text{Ca}^{2+}]_i$. Spontaneous firing of action potentials⁶ evoked similar $[\text{Ca}^{2+}]_i$ transients

Table 1 Characteristics of spontaneous $[\text{Ca}^{2+}]_i$ fluctuations in GH₃B₆ cells

	$[\text{Ca}^{2+}]_i$ mean ($n = 18$)
Prespike	154 (140-168*) nM
Spike maximum	606 (560-654*) nM

* 95% confidence limits

Spiking frequency 7.4 ± 0.6 per min ($\pm$ s.e.m., $n = 23$).

(Fig. 1c, d). Note that the monitoring of the whole cell implies that the intracytosolic diffusion of Ca^{2+} entering during the action potential has to be complete to reach the maximal fura 2 signal. Thus, at times between the onset of the action potentials and the maximum elevation of $[\text{Ca}^{2+}]_i$, changes in $[\text{Ca}^{2+}]_i$ occurring locally (such as those close to the plasma membrane) could be considerably larger than those measured over the entire cell. A comparison of the speed of onset of the $[\text{Ca}^{2+}]_i$ transient with the duration of the action potential (Fig. 1c, d) reveals that intracellular diffusion of Ca^{2+} is relatively slow.

Using the Ca^{2+} -sensitive dye arsenazoIII in large molluscan neurons, $[\text{Ca}^{2+}]_i$ transients associated with the firing of bursts of action potentials or with artificially prolonged single action potentials have been described^{7,8}; in the latter case, $[\text{Ca}^{2+}]_i$ spikes were observed with a similar kinetic appearance to those reported here. This shows that the kinetics of the changes induced in $[\text{Ca}^{2+}]_i$ by the action potential are not affected by the nature of the Ca^{2+} probe (fura 2 or arsenazoIII).

The typical kinetic aspect of the $[\text{Ca}^{2+}]_i$ transient, as observed in the combined experiments shown in Fig. 1, was also seen in spontaneous fluctuations of $[\text{Ca}^{2+}]_i$ detected by single cell monitoring of fura 2 loaded GH₃B₆ cells not subjected to patch clamping. Such oscillations elevated $[\text{Ca}^{2+}]_i$ at regular intervals from a constant prespike level to short-lived peaks (Fig. 2a). The changes in fluorescence at the two excitation wavelengths were perfectly antiparallel, thus a rise in F_{350} occurred simultaneously with a lowering of F_{380} (Figs 2b, c), so the change in fluorescence ratio is certainly a result of altered $[\text{Ca}^{2+}]_i$. The detailed kinetics of the $[\text{Ca}^{2+}]_i$ spike shown in Fig. 2d are indistinguishable from those observed in the combined experiment (Fig. 1c), implying that these $[\text{Ca}^{2+}]_i$ spikes result from enhanced Ca^{2+} influx during action potentials occurring spontaneously in these cells⁶. Indeed, $[\text{Ca}^{2+}]_i$ oscillations stopped when Ca^{2+} was removed from the extracellular medium (data not shown). The transient depolarization during the action potential causes enhanced Ca^{2+} influx via voltage-dependent Ca^{2+} channels^{9,10}. Nifedipine, a dihydropyridine antagonist of the L-channel¹¹ abolished the spiking of $[\text{Ca}^{2+}]_i$ in 8 of the 12 cells tested (Fig. 2d). The failure of nifedipine to affect $[\text{Ca}^{2+}]_i$ transients in some cells could reflect the insensitivity to dihy-

Fig. 1 Action potential induced alterations in $[Ca^{2+}]_i$ in GH_3B_6 cells. $[Ca^{2+}]_i$ was monitored with the fluorescent probe fura 2 during electrophysiological recordings obtained with the patch clamp technique in its whole cell configuration. Action potentials occurred spontaneously (*c, d*) or were elicited by the injection of a depolarizing current (*a, b*). *a*, Ratio of fura 2 fluorescence, measured with a time resolution of 200 ms at excitation wavelengths 350 nm and 380 nm, (F_{350}/F_{380}) of a single patched GH_3B_6 cell; *b*, coincident recording of membrane potential (m.p.) of the same cell maintained at zero current, except for the injection of depolarizing pulses of current (lower trace). *c*, Detailed kinetics of a single $[Ca^{2+}]_i$ transient, following the firing of a spontaneous action potential; *d*, coincident m.p. at zero current.

Methods. GH_3B_6 cells, a subclone of GH_3 cells²² were cultured on glass coverslips for 3–5 days, and loaded with fura 2 either passively from the patch pipette or by exposure to 2 μ M fura 2/AM (Figs 2 and 3, Table 1) in Krebs–Ringer, bicarbonate, HEPES, KRBH (ref. 16) with 0.1% BSA for 30 min at 37 °C. Dual excitation microfluorimetry was performed essentially as described previously^{4,23} on an inverted microscope (epifluorescence mode, objective NIKON F100). Excitation light alternated rapidly (50 Hz) between $\lambda_1 = 350$ nm and $\lambda_2 = 380$ nm (SPEX fluorimeter, Glen Creston, London). Fluorescence of single cells, superfused with KRBH at 37 °C, was monitored through a circular diaphragm slightly larger than the cells and an interference filter ($\lambda = 500$ nm) by photon counting. From the ratio $R = F_{350}/F_{380}$, $[Ca^{2+}]_i$ was calculated using the formula:

$$[Ca^{2+}]_i = K_d \cdot \beta (R - R_{min}) / (R_{max} - R)$$

where K_d is 225 nM (ref. 3); R_{min} , R_{max} , and β were determined in separate experiments, using the ionophore ionomycin to equilibrate $[Ca^{2+}]_i$ with ambient $[Ca^{2+}]$, kept at either below 10 nM (R_{min}) or at 5 mM (R_{max}); $\beta = F_{380}(min)/F_{380}(max)$ was obtained during a change between these conditions, correcting for eventual loss of fura 2 (ref. 23). R_{min} , R_{max} and β averaged 0.42 ± 0.03 ($\pm$ s.e.m., $n = 16$), 4.3 ± 0.3 ($n = 24$) and 4.7 ± 0.3 ($n = 17$) respectively, and were distinct from the corresponding constants obtained with calibration solutions²³. Combined microfluorimetric and electrophysiological experiments were performed at room temperature using the whole cell recording configuration of the patch clamp technique², as described previously²⁴; blunt whole cell patch pipettes (resistance 3–10 megohms) contained the following solution (in mM) K^+ gluconate (140), $MgCl_2$ (1), EGTA (1.1), HEPES (5), fura 2 (0.03), pH 7.1. Extracellular medium was (in mM): NaCl (150), KCl (5), $CaCl_2$ (10), $MgCl_2$ (2), glucose (5), HEPES (5), pH 7.3.

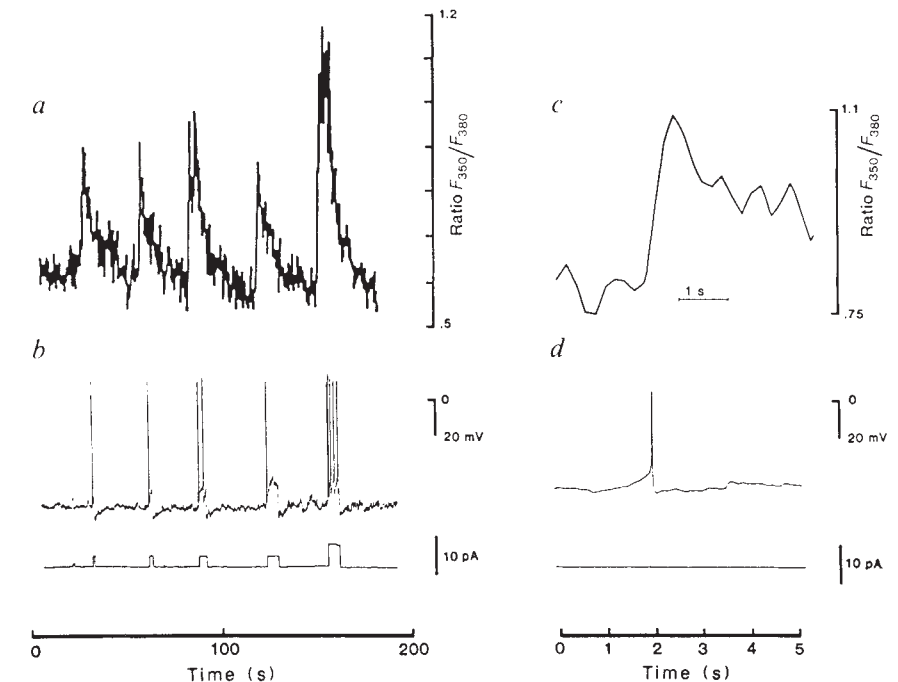
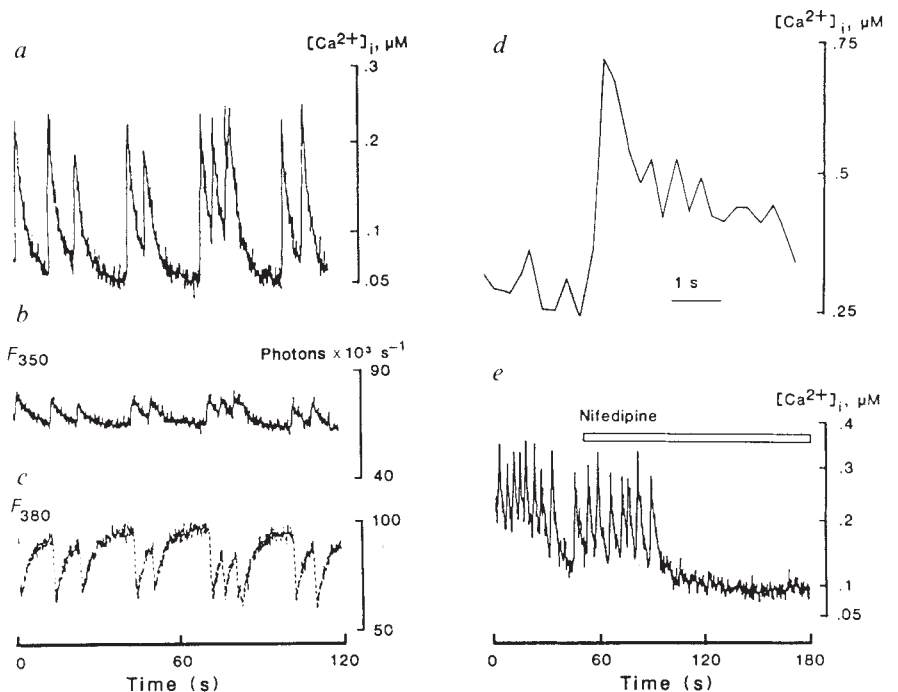


Fig. 2 Spontaneous oscillations of $[Ca^{2+}]_i$ in GH_3B_6 cells. *b* and *c*, Fluorescence of a single fura 2 loaded cell, monitored at excitation wavelengths 350 nm (F_{350} , *b*) and 380 nm (F_{380} , *c*) simultaneously, with a time resolution of 100 ms; *a*, ratio of fura 2 fluorescence (F_{350}/F_{380}) from the same experiment, calibrated for $[Ca^{2+}]_i$. *d*, Kinetics of a single $[Ca^{2+}]_i$ transient. Ratio of fura 2 fluorescence, calibrated for $[Ca^{2+}]_i$ in a single GH_3B_6 cell monitored at 200 ms time resolution (unfiltered data). The time for the transition from prespike $[Ca^{2+}]_i$ to the maximal level (in this example 200 ms) averages 376 ± 17 ms ($\pm$ s.e.m., $n = 56$). *e*, The Ca^{2+} channel blocker nifedipine abolishes $[Ca^{2+}]_i$ spiking. Ratio of fura 2 fluorescence, calibrated for $[Ca^{2+}]_i$, monitored from a single GH_3B_6 cell. The bar indicates the time during which nifedipine was present at 10 μ M.



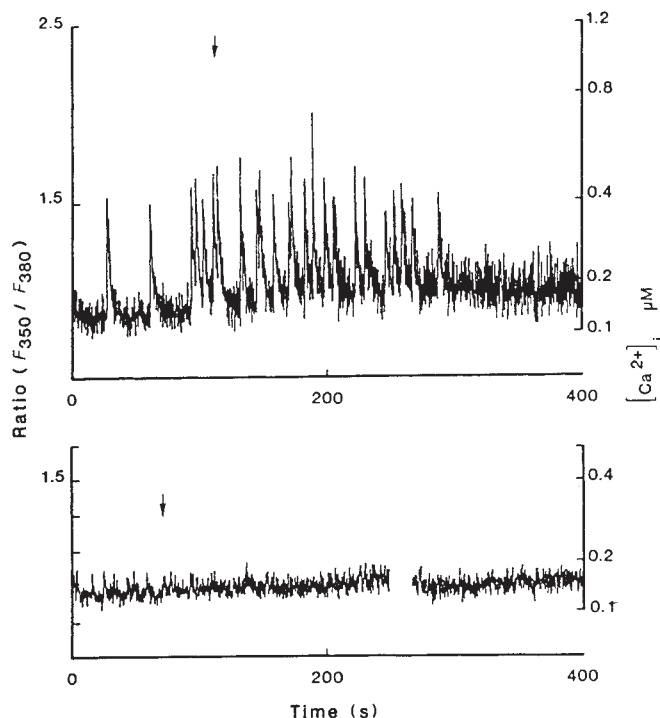


Fig. 3 Somatostatin abolishes spontaneous $[Ca^{2+}]_i$ transients. Ratio of fura 2 fluorescence, calibrated for $[Ca^{2+}]_i$, monitored with a time resolution of 100 ms in a spontaneously spiking GH₃B₆ cell (upper panel) and a quiescent cell (lower panel). The arrow indicates the time at which the infusion of somatostatin (0.1 μ M final concentration) started.

dropyridine antagonists of certain states of the L-channel and/or the presence in GH₃B₆ cells of other Ca^{2+} channel types^{1,10-12}.

The data in Table 1 indicate that the fluctuations in $[Ca^{2+}]_i$ arising from spontaneous action potentials permit Ca^{2+} -dependent cell activation. The dependence on $[Ca^{2+}]_i$ for prolactin secretion was determined in permeabilized GH₃ cells; stimulation occurs over the range 0.2 μ M–10 μ M free $[Ca^{2+}]$ (ref. 13). The average rise in $[Ca^{2+}]_i$ associated with spontaneous action potentials described here is therefore sufficient to trigger prolactin secretion; in fact, the mean maximal $[Ca^{2+}]_i$ (606 nM) is close to the $[Ca^{2+}]$ required for half-maximal stimulation of secretion in the absence of any other stimulus¹³. Thus, the spontaneous $[Ca^{2+}]_i$ transients are likely to cause basal prolactin secretion.

Somatostatin, a hypothalamic peptide inhibitor of growth hormone secretion⁵ lowers basal prolactin secretion in GH₃ cells^{14,15}. In Fig. 3 it is shown that somatostatin gradually reduces spontaneous spiking activity. In contrast, somatostatin did not significantly affect $[Ca^{2+}]_i$ in GH₃B₆ cells which did not display spontaneous $[Ca^{2+}]_i$ fluctuations. Study of $[Ca^{2+}]_i$ in a population of cells monitored previously showed that somatostatin reduced 'basal' $[Ca^{2+}]_i$ (ref. 16). Our current data now explain this effect: the apparently stable 'basal' $[Ca^{2+}]_i$ includes a large number of random $[Ca^{2+}]_i$ transients not visible when monitoring a cell population; their abolition results in a slight reduction of the average $[Ca^{2+}]_i$ (ref. 16). Also, the abolition of $[Ca^{2+}]_i$ transients by somatostatin causes the inhibition of basal secretory activity, which in turn is due to spontaneous $[Ca^{2+}]_i$ spiking. Results obtained recently with normal pituitary cells (manuscript in preparation) indicate that somatostatin acts similarly on its physiological target.

Because of the large ion gradients across the plasma membrane, cellular Ca^{2+} homeostasis requires energy. The most efficient way to trigger Ca^{2+} -dependent intracellular processes therefore is by short transients, whereas prolonged increases in $[Ca^{2+}]_i$ would consume substantial amounts of energy. Oscilla-

tions in $[Ca^{2+}]_i$ have been observed in association with cytoplasmic streaming¹⁷, cell division^{18,19}, and lymphocyte activation¹⁸. Furthermore, repetitive $[Ca^{2+}]_i$ transients are also generated in non-excitable cells by hormones which stimulate the production of the calcium-mobilizing second messenger inositol-1,4,5-trisphosphate^{20,21}. The principle of regulating Ca^{2+} -dependent intracellular processes by variation of the frequency of short $[Ca^{2+}]_i$ transients could apply generally in signal transduction.

The present study demonstrates that the short (5–10 msec) concerted opening of voltage-dependent Ca^{2+} channels during the action potential is effective in creating a marked transient elevation of $[Ca^{2+}]_i$. This implies that the secretory activity of excitable endocrine cells is indeed regulated via the well-documented^{1,9} influence of hormones, neurotransmitters and nutrients on electrical activity.

The authors thank Dr Roger Y. Tsien, and Dr Tullio Pozzan for their support. This work was funded by Fonds National Suisse, CNRS France, European Science Foundation and the Fondation Carlos and Elsie de Reuter.

Received 19 June; accepted 8 September 1987.

- Ozawa, S. & Sand, O. *Physiol. Rev.* **66**, 887–952 (1986).
- Hamill, O. P. *et al. Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
- Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440–3450 (1985).
- Tsien, R. Y., Rink, T. J. & Poenie, M. *Cell Calcium* **6**, 145–157 (1985).
- Brazeau, P. *et al. Science* **179**, 77–79 (1973).
- Kidokoro, Y. *Nature* **258**, 741–742 (1975).
- Thomas, M. V. & Gorman, A. L. F. *Science* **196**, 531–533 (1977).
- Gorman, A. L. F. & Thomas, M. V. *J. Physiol.* **275**, 357–376 (1978).
- Taraskevich, P. S. & Douglas, W. W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4064–4067 (1977).
- Reuter, H. *Nature* **301**, 569–574 (1983).
- McCleskey, E. W. *et al. J. exp. Biol.* **124**, 177–190 (1986).
- Armstrong, C. M. & Matteson, D. R. *Science* **227**, 65–66 (1985).
- Ronning, S. A., & Martin, F. J. *J. biol. Chem.* **261**, 7840–7845 (1986).
- Schonbrunn, A., & Tashjian, A. H. Jr *J. biol. Chem.* **253**, 6473–6478 (1978).
- Dorflinger, L. J. & Schonbrunn, A. *Endocrinology* **113**, 1551–1558 (1983).
- Schlegel, W. *et al. Cell Calcium* **5**, 223–236 (1984).
- Ridgway, E. B. & Durham, A. C. H. *J. cell Biol.* **69**, 223–226 (1976).
- Tsien, R. Y. & Poenie, M. *Trends biochem. Sci.* **11**, 450–455 (1986).
- Poenie, M. *et al. Science* **233**, 886–889 (1986).
- Woods, N. M., Cuthbertson, K. S. R. & Cobbold, P. H. *Nature* **319**, 600–602 (1986).
- Woods, N. M., Cuthbertson, K. S. R. & Cobbold, P. H. *Cell Calcium* **8**, 79–100 (1987).
- Tashjian, A. H. Jr, *Meth. Enzym.* **58**, 527–535 (1979).
- Schlegel, W. *et al. J. Receptor Res.* (in the press).
- Duffy, B., McDermott, A. & Barker, J. L. *Biochem. biophys. Res. Commun.* **137**, 388–396 (1986).

At least two genes reside within a large intron of the *dunce* gene of *Drosophila*

Chun-nan Chen*, Tom Malone†, Steven K. Beckendorf† & Ronald L. Davis*‡

* Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824, USA

† Department of Molecular Biology, University of California, Berkeley, California 94720, USA

The *dunce* locus of *Drosophila melanogaster* is considered to house a gene involved in memory, because flies carrying lesions at the locus have shortened memory of several different conditioned behaviours^{1–6}. Our recent partial characterization of the gene at the molecular level, along with prior genetic and biochemical evidence, recently provides compelling evidence that the gene codes for the enzyme cAMP phosphodiesterase⁷. The observation that the gene encodes at least six overlapping poly(A)⁺ RNA molecules ranging in size from 4.2 to 9.5 kilobases (kb) (ref. 8), suggests that the gene is extraordinarily complex. Here we provide the sequence of a *dunce* complementary DNA clone and the corresponding genomic coding regions which show that the organization of the gene is elaborate. The cDNA clone defines *dunce* exons which are separated by a large intron of 79 kb. More importantly, at least

‡ To whom correspondence should be addressed at Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA.

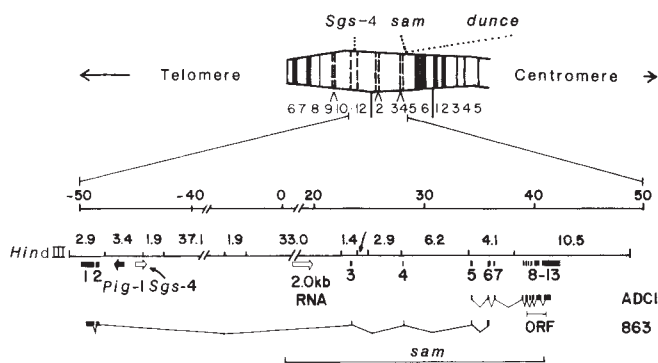


Fig. 1 Diagram of the 3C6-3E5 chromosomal interval showing *Sgs-4* at chromomere 3C11-12 (ref. 12); and *dunce* and the complementation group *sam*, both of which have been mapped cytogenetically to chromomere 3D4 (ref. 18). The expanded view of the genomic DNA in the 3C11-3D4 interval is mapped in *Hind*III fragments relative to an arbitrary coordinate system from -50 to 50 (ref. 9). Each unit of the coordinate system represents 1 kb. The numbering does not include a 7.3-kb insertion element between coordinates 2 and 5 (ref. 9). Positions of *dunce* exons (numbered 1-13) and other genes are shown below the restriction map. The *dunce* exons are defined by the cDNA clone, ADC1, previously isolated from an oligo(dT)-primed cDNA library⁷; and clone 863. ADC1 contains the open reading frame (ORF) for the cAMP phosphodiesterase. The limits of *sam* are shown.

Methods. A primer-extension cDNA library using *Drosophila* adult poly(A)⁺ RNA and an RNA-complementary oligonucleotide in exon 6 was constructed in the vector λ gt10 essentially as described²⁷. The library was screened with cloned probes representing exons 3 and 5. Exon 3 was defined before the library screening by *S*₁ nuclease experiments and RNA blot analysis (C.-N.C. and R.L.D., manuscript in preparation).

two other genes are shown to reside within the large intron, including the well-defined glue protein gene, *Sgs-4*. The location of *dunce* exons relative to the molecular breakpoints of chromosomal aberrations with defined cytological positions indicates that the *dunce* gene extends over more than five polytene chromosome bands.

The *dunce* locus was isolated as overlapping clones representing genomic segments extending rightward (centromere proximally; see Fig. 1) from the nearby gene, *Sgs-4* (ref. 9). We identified a genomic region of ~25 kb (coordinates 21-46 in Fig. 1) as containing exons of the *dunce* RNA molecules by RNA blotting experiments⁸, isolated cDNA clones from oligo(dT)-primed cDNA libraries⁷, and sequenced the cDNA clones and nearly the complete genomic region from coordinate 21 to coordinate 46 (ref. 7; C.-N.C. and R.L.D., manuscript in preparation). The cDNA clone ADC1 (Fig. 1) defines exons 5 to 13, as well as the open reading frame which encodes the cAMP phosphodiesterase enzyme. But because the longest cDNA clone previously isolated was only 2.2 kb, and the smallest *dunce* poly(A)⁺ RNA molecule is 4.2 kb, we sought cDNA clones representing more of the 5' sequence information of *dunce* RNA molecules.

To isolate cDNA copies of the 5' regions of the *dunce* RNA molecules, we constructed a primer-extension cDNA library (Fig. 1). Eighteen cDNA clones representing *dunce* RNAs were recovered from this library and analysed by restriction mapping, sequence analysis, and hybridization experiments to genomic clones. One clone, with a unique structure and named 863, is described in detail here. The sequence of 863 defines exons 3, 4, 5, and part of 6 as judged by comparison with the genomic sequence. But 863 contains an additional 1.1 kb of sequence information not found in the genomic region from coordinate 21 to coordinate 46, suggesting that there are other exons to the left of exon 3. Approximately 80 kb of genomic sequence contained in clones to the left of exon 3 was surveyed by hybridiz-

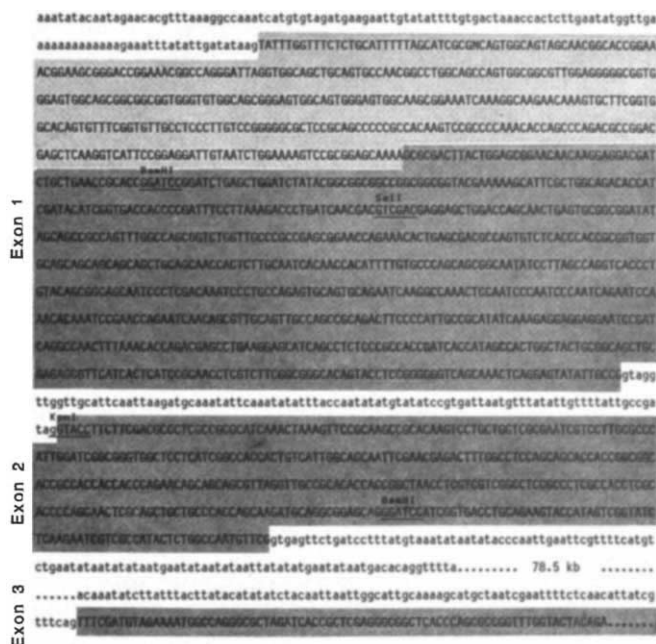


Fig. 2 Nucleotide sequence of exons 1, 2 and a portion of 3 aligned with the corresponding genomic sequences (Fig. 1). Dark grey shading, sequences found in 863. The terminal nucleotide in 863 is not the 5' end of exon 1 as *S*₁ experiments have shown that the exon continues to a probable splice-site 381 residues more 5' to the end of 863. This region is shaded light grey. Primer-extension experiments also show that at least one *dunce* RNA contains ~500 more nucleotides coded for by exons upstream of exon 1 (C.-N.C. and R.L.D., manuscript in preparation). Sequence analysis was by cloning genomic or cDNA fragments into M13 vectors with subsequent dideoxy sequencing.

ation with 863. Only the 2.9-kb *Hind*III fragment at coordinate -48 to coordinate -51 was found to hybridize. A major portion of this 2.9-kb genomic fragment was sequenced and aligned with the sequence in 863 (Fig. 2). The alignment reveals two additional exons, designated 1 and 2, which are separated by a small intron of 98 base pairs (bp). In contrast, exons 2 and 3 are separated by a very large intron of about 79 kb. All the intron-exon junctions conform to the consensus sequence of splice sites, indicating that 863 represents a portion of an authentic *dunce* RNA molecule and not a cDNA clone with an artefactual arrangement.

RNA blotting experiments confirm the authenticity of 863 and assign exons 1 and 2 as coding for the 9.5-kb RNA and at least one of about 7.2 kb. Single-stranded probes representing the RNA-complementary sequence or the RNA-like sequence of exons 1, 2 or 6 were used to probe blots of the poly(A)⁺ RNA population of adult flies (Fig. 3). A probe representing the RNA-complementary sequence of exon 6 hybridizes to all molecules previously assigned as *dunce* RNAs (ref. 8), showing that each *dunce* RNA uses this exon. In contrast, probes representing the RNA-complementary sequence of exons 1 and 2 hybridize to the 9.5- and 7.2-kb RNAs, showing that only a subset of the *dunce* poly(A)⁺ RNAs use these exons. RNAs other than 9.5 and 7.2 kb are also detected by probes of exons 1 and 2, but we have little information about these molecules.

A surprising picture emerges when the structure of *dunce* as defined by 863 is superimposed upon the known location of other genes in the region. The *Sgs-4* gene was previously assigned to coordinate -44 to coordinate -45 in Fig. 1 (refs 10-12). This places *Sgs-4* in the 79-kb intron of the *dunce* gene. *Sgs-4* is a well-characterized gene which encodes a protein component of the glue produced in the larval salivary glands¹⁰⁻¹⁷. It is expressed primarily in the glands of late third instar larvae and is transcribed in the same direction as *dunce*. The transcriptional unit of *Sgs-4* has been rigorously defined

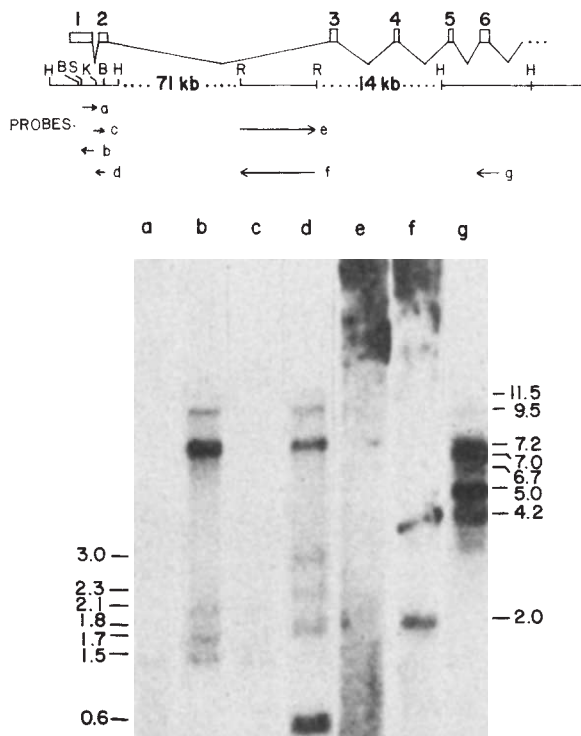


Fig. 3 Analysis of RNAs from the *dunce* region. Top, diagram of *dunce* exons 1 to 6 aligned on a partial restriction map. Compare with Fig. 2 for the locations of restriction sites in the sequence of exons 1 and 2. Single-stranded probes a–g were used to probe blots of 10 μ g of adult poly(A)⁺ RNA after fractionation by denaturing gel electrophoresis⁸. The size of the *dunce* RNAs had been determined⁸ as ranging from 4.5 to 9.6 kb and that of the RNA from coordinates 17–20 as 1.8 kb relative to *Escherichia coli* ribosomal RNA standards. The revised sizes indicated here were obtained by comparison with the BRL RNA ladder. Probes were synthesized from fragments subcloned into M13 vectors essentially as described²⁸ and isolated from denaturing polyacrylamide or agarose gels. Probes a and b represent opposite strands of the *Sall* (S)–*KpnI* (K) fragment, and probes c and d the strands of the *KpnI*–*BamHI* (B) fragment. The hybridization to the 9.5-kb RNA in lane g and the 11.5-kb RNA in lane d is very weak.

with its promoter at coordinate –45.6 (ref. 10). All the *cis*-acting elements necessary for a high level of expression, dosage compensation, and normal spatial and temporal specificity have been defined in the sequence between 840 bp 5' and 130 bp 3' of the transcription unit¹⁷. In addition, a partially characterized gene called *Pig-1* (preintermolt gene-1) is just distal to *Sgs-4* (P. Mathers and E. Meyerowitz, personal communication; see ref. 17), and therefore also is within the large intron of *dunce*. Like *Sgs-4*, *Pig-1* is expressed primarily in larval salivary glands, but its expression begins earlier and peaks in the second, rather than the third, instar (Fig. 4). The *Pig-1* transcriptional unit is about 630 bp long and lies just 840 bp distal to *Sgs-4* between coordinates –46.4 and –47.0 (B. Rogers and S.K.B., unpublished). It encodes a poly(A)⁺ RNA of ~750 bp which is transcribed from the opposite strand as *dunce* or *Sgs-4* (Fig. 4). Interestingly, neither of these two *dunce*-intronic genes has introns of its own (ref. 10; B. Rogers and S.K.B., unpublished). This may be significant because *Sgs-4* is the only one of the five characterized *Sgs* genes which lacks introns.

There is suggestive evidence that there may be other genes in the 79-kb intron of *dunce*. For example, a single-stranded probe representing a unique sequence at coordinates 17–20 hybridizes to a 2.0-kb transcript in adult flies (Fig. 3). This transcript is encoded by the same strand as *dunce* but shows a distinct developmental expression profile, existing at detectable levels only at the third larval instar and later⁸. In contrast, most of

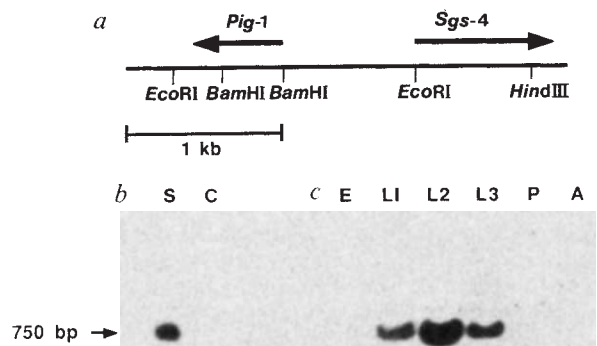


Fig. 4 a, Location of the *Pig-1* gene and b, c, its pattern of spatial and developmental expression. a, Diagram of the *Pig-1* and *Sgs-4* genes aligned on a partial restriction map. b, Tissue specificity of *Pig-1* transcription. A single-stranded RNA extending in the 5' to 3' direction and representing a unique genomic sequence from the left *EcoRI* site to the right *BamHI* site was used to probe a blot of RNA from dissected salivary glands (S) or from larval carcasses remaining after the salivary glands had been removed (C). To confirm the orientation of the *Pig-1* gene, the opposite strand was used to probe a similar blot. No hybridization was detected in this case. c, Temporal regulation of *Pig-1*. A nick-translated probe containing the same *EcoRI*–*BamHI*–*BamHI* sequence was used to probe a blot of RNA from embryos (E), first, second, and third instar larvae (L1, L2, L3), pupae (P), and adults (A). Each lane contained 32 μ g of total RNA.

the *dunce* RNAs reach detectable levels during embryogenesis or the first larval instar⁸. Therefore, this RNA seems to be encoded by a transcriptional unit distinct from *dunce*. Moreover, a genetic complementation group called *sam* (sperm-amotile) can be placed close to or within the large intron of *dunce*. Mutations which define this complementation group cause male sterility and have been placed genetically to the right of the right breakpoint of the chromosome *Df(1)N^{64j15}* (ref. 8), for which the molecular breakpoint is between coordinate –2 and coordinate 0 (ref. 9). The right limit of *sam* is the open reading frame of *dunce*, because *sam* mutations map genetically to the left of *dunce* mutants¹⁹ which alter the *K_m* or thermostability of the cAMP phosphodiesterase^{20,21}, and therefore must be in the protein coding region. It is possible that the 2.0-kb transcript from coordinates 17–20 corresponds to the *sam* complementation group.

Previously mapped chromosomal breakpoints allow a minimum estimate of the chromosomal domain occupied by *dunce*. *Sgs-4* has been placed in chromomeres 3C11–12 by its molecular location relative to that of chromosomal aberrations which break in this region¹². Similarly, exons 3–13 of *dunce* are between chromosomal breakpoints on either side of chromomere 3D4 (refs 7–9). Therefore, *dunce* must extend over a minimum of five polytene chromosome bands.

Several other eukaryotic genes are known to have unusual, but less elaborate organizations. For example, adjacent genes which overlap at their 3' ends have been described in the mouse²² and in *Drosophila*²³. A cuticle protein gene has been found to be embedded within an intron of the *Gart* gene of *Drosophila*²⁴. More recently, a gene was discovered which is encoded from the opposite strand of the gonadotropin-releasing hormone gene of the rat²⁵. One might argue that genes with large introns, such as *dunce*, are likely to contain biologically important information in the introns. However, the *Antennapedia* gene of *Drosophila* contains several introns as large as 60 kb, but no other genes have been found within these²⁶. The structural organization exemplified by *dunce* poses several questions about the regulation and evolution of genes and the processing of their primary transcripts. First, is there any relationship between the expression of *dunce* and the genes within *dunce*, or are the expression patterns completely independent of one another?

Second, how has this organization evolved? Perhaps the genes in the large intron have transposed between *dunce* exons, or alternatively, exons 1 and 2 of *dunce* may have recently been recruited to be part of *dunce*. Third, is *dunce* expressed as a primary transcript of >100 kb with a 79-kb intron subsequently removed, or is it more reasonable to invoke *trans*-splicing, discontinuous transcription, or other more novel processes to accomplish the joining of exons 2 and 3?

We thank P. Mathers and E. Meyerowitz for alerting us to the location of *Pig-1*. This work was supported by an NIH grant, a Research Career Development Award and an award from the McKnight Foundation to R.L.D.; and by an NIH grant to S.K.B.

Received 2 July; accepted 7 September 1987.

1. Dudai, Y., Jan, Y., Byers, D., Quinn, W. G. & Benzer, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1684-1688 (1976).
2. Dudai, Y. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5445-5448 (1983).
3. Tempel, B. L., Bonini, N., Dawson, D. & Quinn, W. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1482-1486 (1983).
4. Tully, T. & Quinn, W. J. *comp. Physiol.* **157**, 263-277 (1985).
5. Kyriacou, C. P. & Hall, J. C. *Nature* **308**, 62-64 (1984).
6. Aceves-Pina, E. O. et al. *Cold Spring Harb. Symp. quant. Biol.* **48**, 831-840 (1984).
7. Chen, C.-N., Denome, S. & Davis, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9313-9317 (1986).
8. Davis, R. & Davidson, N. *Molec. cell. Biol.* **6**, 1464-1470 (1986).
9. Davis, R. & Davidson, N. *Molec. cell. Biol.* **4**, 358-367 (1984).
10. Muskavitch, M. A. & Hogness, D. S. *Cell* **29**, 1041-1051 (1982).
11. Muskavitch, M. A. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7362-7366 (1980).
12. McGinnis, W., Farrell, J. Jr & Beckendorf, S. K. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7367-7371 (1980).
13. McGinnis, W., Shermoen, A. W. & Beckendorf, S. K. *Cell* **34**, 75-84 (1983).
14. Shermoen, A. W. & Beckendorf, S. K. *Cell* **29**, 601-607 (1982).
15. McGinnis, W., Shermoen, A. W., Heemskerk, J. & Beckendorf, S. K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1063-1067 (1983).
16. Shermoen, A. W., Jongens, J., Barnett, S. W., Flynn, K. & Beckendorf, S. K. *EMBO J.* **6**, 207-214 (1987).
17. McNabb, S. L. & Beckendorf, S. K. *EMBO J.* **5**, 2331-2340 (1986).
18. Salz, H. K., Davis, R. L. & Kiger, J. A. *Genetics* **100**, 587-596 (1982).
19. Salz, H. K. & Kiger, J. A. *Genetics* **108**, 377-392 (1984).
20. Kauvar, L. M. *J. Neurosci.* **2**, 1347-1358 (1982).
21. Davis, R. L. & Kauvar, L. M. *Adv. cyclic Nucleotide Res.* **16**, 393-402 (1984).
22. Williams, T. & Fried, M. *Nature* **332**, 275-281 (1986).
23. Spencer, C., Gietz, D. & Hodgetts, R. *Nature* **322**, 279-281 (1986).
24. Henikoff, S., Keene, M., Fichtel, K. & Fristrom, J. *Cell* **44**, 33-42 (1986).
25. Adelman, J. P., Bond, C. T., Douglass, J. & Herbert, E. *Science* **235**, 1514-1517 (1986).
26. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. J. *EMBO J.* **5**, 733-739 (1986).
27. Toole, J. et al. *Nature* **312**, 342-347 (1984).
28. Burke, J. F. *Gene* **30**, 63-68 (1984).

Latent synaptic pathways revealed after tetanic stimulation in the hippocampus

Richard Miles & Robert K. S. Wong

Department of Neurology, 4-408 P and S, Columbia University, 630 West 168th Street, New York, New York 10032, USA

Synaptic plasticity may result from changes at existing synapses or from alterations in the number of functional synaptic connections¹. In the hippocampus excitatory synaptic strength is persistently enhanced after tetanic stimulation²⁻⁵. Here we report that latent synaptic pathways may also become functional. Simultaneous intracellular recordings were made from pairs of CA3 pyramidal cells in slices from guinea pig hippocampus. After stimulating afferent fibres repetitively, polysynaptic excitatory pathways between previously unconnected cells became apparent. The efficacy of recurrent inhibitory circuits was also reduced. The loss of inhibitory control is of interest because latent excitatory pathways are revealed after pharmacological suppression of inhibition. This plasticity in local synaptic circuits leads to the emergence of synchronous firing in groups of CA3 cells. The formation of groups of associated cells and the ability of some cells to initiate synchronous firing in a larger cell group through recurrent pathways is reminiscent of several models of information storage and recall in the cortex⁶⁻¹⁰.

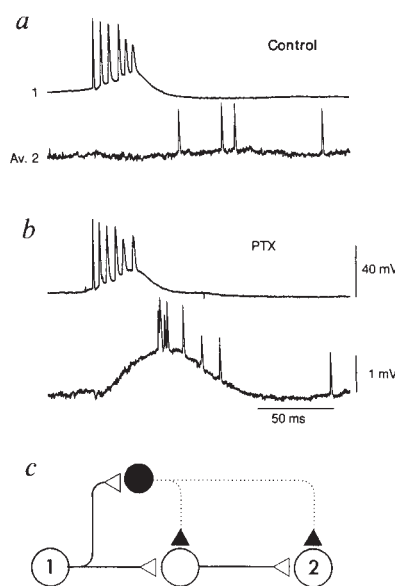


Fig. 1 Polysynaptic excitation revealed after suppressing inhibition. *a*, No excitatory interaction was detected in averages of responses in cell 2 to bursts of 3-7 action potentials in cell 1 ($n = 52$; Av., average). *b*, An excitatory interaction was apparent in averaged responses ($n = 52$) made 21-23 min after exposure to 5 μ M picrotoxin (PTX). E.p.s.ps were evoked by 27 of 52 presynaptic bursts. Their latency, measured from the first presynaptic spike was 7 ± 3 ms (mean $\pm$ s.d.). Their time to peak was 27 ± 9 ms and their amplitude was 1.9 ± 0.6 mV at potentials between -64 and -67 mV. Bursts of 3-7 action potentials in cell 1 were evoked by depolarizing current injections at 0.5 Hz. Small spikes in averaged postsynaptic traces reflect action potentials. *c*, Diagram of local collateral synaptic circuitry which may be responsible. Triangles, synapses; open circles, pyramidal cells, with recordings made from cells 1 and 2. Filled circle, inhibitory cell which could act on the postsynaptic or the intercalated cell to block transmission.

Methods. Slices (400 μ m) were prepared from guinea-pig hippocampus. They were cut either transversely, or longitudinally in a plane orthogonal to the CA2-3a pyramidal cell layer. They were maintained in an interface chamber at 37 $^{\circ}$ C, at pH 7.4, exposed to 5% CO_2 in O_2 and perfused with a solution containing (in mM) Na^+ 150, K^+ 4.5, Ca^{2+} 1.5-2.5, Mg^{2+} 1-1.5, Cl^- 130, HCO_3^- 25 and D-glucose 10 (in mM). Intracellular recordings were made from cells within about 400 μ m of each other in CA3 stratum pyramidale with electrodes containing 2 M potassium acetate. Extracellular recording electrodes contained 1 M NaCl. Bipolar tungsten electrodes were used to stimulate either mossy fibres or the longitudinal pathway in the stratum radiatum. Synaptic interactions were assessed by digitally averaging 30-60 responses of one cell to firing in another cell, occurring spontaneously or induced by depolarizing current injections at 0.5 Hz. The putative post-synaptic cell was usually hyperpolarized to prevent firing. Capacitive electrode coupling artefacts were digitally suppressed.

Local axon collaterals of CA3 pyramidal cells form excitatory synapses with inhibitory cells^{11,12} and other pyramidal cells^{13,14}. Synapses between pyramidal cells are sufficiently strong for firing in a pre-synaptic cell to evoke post-synaptic firing. Both repetitive synaptic activation by bursts of pre-synaptic action potentials and a low post-synaptic firing threshold contribute to this unusually effective synaptic coupling¹³⁻¹⁵. With strong recurrent synapses between pyramidal cells, firing in one neuron might sequentially excite a chain of synaptically connected cells. Indeed activation of one neuron evoked polysynaptic excitatory postsynaptic potentials (e.p.s.ps) in another cell in 45 of 270 dual recordings from the CA3 region, when synaptic inhibition was suppressed with the γ -aminobutyric acid antagonist picrotoxin (5-100 μ M). E.p.s.p. latency was >4 ms and transmission failed more often than at monosynaptic connections^{14,15}.

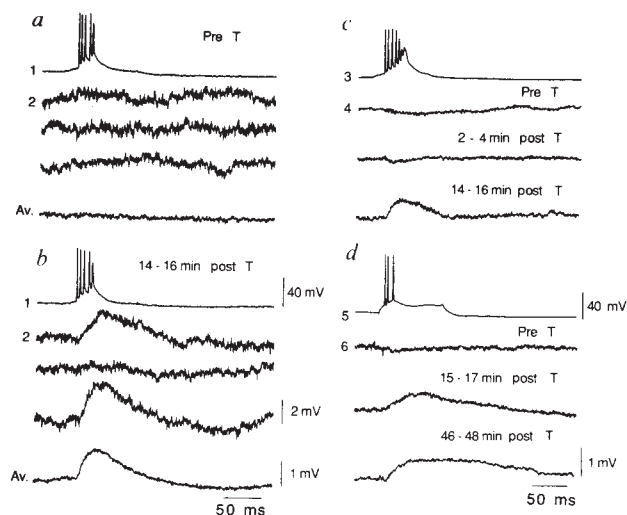


Fig. 2 Latent polysynaptic e.p.s.ps revealed after tetanic stimuli. *a*, Three responses and the average of 50 responses in cell 2 initiated by bursts evoked in cell 1 before tetanic stimulation (pre T). *b*, Single responses and an average (post T) obtained 14–16 min after tetanic stimulation (20 Hz, 10 s, $\times 2$) of mossy fibres in a transverse slice. E.p.s.ps were evoked by 32 of 50 pre-synaptic bursts. Their mean latency from the first presynaptic spike was 6 ± 3 ms and their mean amplitude was 2.4 ± 0.6 mV at potentials between -68 and -72 mV. *c*, New connections were not revealed immediately after stimulation. Averages of responses from cell 3 elicited by stimulation of cell 4, obtained before (pre T), and at 2–4 min and 14–16 min after, stimulating longitudinal fibres (50 Hz, 5 s, $\times 2$). *d*, Recurrent excitatory pathways were strengthened by further tetani. Averages ($n = 45$ responses) from cells 5 and 6 before stimulation (pre T) and at 15–17 min after (post T) mossy fibre tetani (50 Hz, 5 s, $\times 2$). The rising phase and later components of the averaged response in cell 6 were enhanced at 46–48 min following further tetanic stimuli at 23 min.

When synaptic inhibition was functional, polysynaptic excitatory interactions were detected in a much smaller proportion, 5 out of 550, cell pairs.

The possibility that synaptic inhibition might suppress excitatory transmission in polysynaptic pathways was examined in recordings made from cell pairs where no excitatory connection was initially apparent. In 8 of 48 interactions, depolarizing events in one cell initiated by firing in another cell were revealed on exposure to picrotoxin (Fig. 1). These events were considered to be mediated by newly functional polysynaptic excitatory pathways¹⁵. E.p.s.ps had a slow waveform (time to peak 15–30 ms, amplitude 2–3 mV) similar to summed e.p.s.ps elicited by bursts of action potentials at monosynaptic recurrent connections¹⁴. However newly functional pathways were not monosynaptic because single action potentials usually evoked no response and responses to presynaptic bursts occurred with long variable latencies (6–30 ms) and with some transmission failures (30–70% of trials). Anatomical changes seem unlikely in the 15–25 min during which connections became apparent. Instead, as in the neuronal system for swimming in *Tritonia*¹⁶, functional excitatory connectivity in the hippocampus seems to depend on the efficacy of synaptic inhibition. Recurrent inhibition may suppress excitatory interactions (Fig. 1) by shunting a weaker dendritic e.p.s.p. on the postsynaptic cell or may act to suppress firing in intercalated cells of the polysynaptic pathway^{12,15}.

We also investigated whether a similar modification of local synaptic circuitry occurs following tetanic stimulation of afferent fibres^{2-5,17-20}. Two pathways were tetanized in different experiments (Fig. 2). Mossy fibres were stimulated in transverse hippocampal slices. In longitudinal slices from the CA3 region the longitudinal association pathway^{21,22} was stimulated. Two tetani

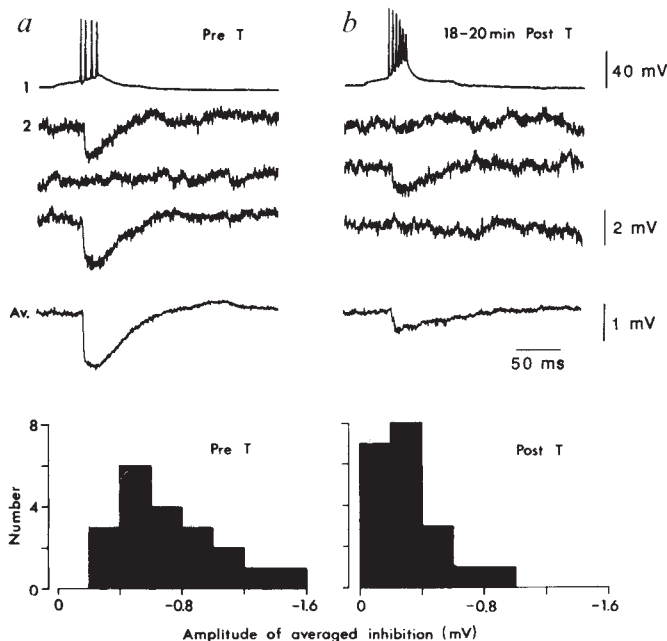


Fig. 3 Recurrent inhibition is reduced after tetanic stimulation. *a*, Single responses and average of 60 responses in cell 2 to bursts of 3–6 action potentials in cell 1 (pre T). I.p.s.ps of time to peak 3–7 ms were evoked in 52 of 60 trials, with mean latency 2.8 ± 0.6 ms from the first spike in cell 1. This was the strongest inhibitory coupling observed in these experiments. *b*, Three responses and the average of 60 responses recorded 18–20 min after tetanic stimulation of mossy fibres (50 Hz, 5 s, $\times 2$). I.p.s.ps of amplitude > 0.5 mV were detected in 16 of 60 trials at similar postsynaptic membrane potentials, -65 to -70 mV. The amplitude of the averaged i.p.s.p. was reduced from 1.4 to 0.5 mV. E.p.s.ps were not seen when the i.p.s.p. failed, suggesting that potentiation of an underlying e.p.s.p. did not contribute to the reduction in inhibitory responses. *c*, Peak amplitude of the averaged synaptic inhibition elicited in 20 cells by 30–60 bursts of pre-synaptic action potentials before tetanic stimulation (pre T) and at 12–20 min after tetanic stimulation (post T). The effects of tetani were tested on five cell pairs with inhibitory coupling from longitudinal slices and 15 cell pairs from transverse slices.

of duration 5–10 s and frequency 20–50 Hz separated by 30 s were given.

Recordings were made from pairs of CA3 cells where no excitatory connection was detected. Communication between cells was tested again 12–20 min after tetanic stimulation. Polysynaptic excitatory interactions were revealed in 5 of 48 cases in transverse slices and in 3 of 22 cases in longitudinal slices (Fig. 2*a, b*). Excitatory pathways were not revealed immediately after afferent stimulation but after a delay of several minutes (Fig. 2*c*). In one case when an excitatory connection was uncovered, recordings were maintained for long enough to show that its efficacy was enhanced further by repeated tetanic stimulation (Fig. 2*d*).

Because latent excitatory connections were revealed when inhibition was pharmacologically suppressed (Fig. 1), it seemed possible that tetanic stimulation might also depress synaptic inhibition¹⁷⁻²⁰. Inhibitory interactions, where one burst-firing CA3 cell evoked inhibitory post-synaptic potentials (i.p.s.ps) in another cell¹², were characterized and re-examined after tetanic stimulation (Fig. 3). These recurrent pathways were presumed to be disynaptic because i.p.s.ps were not elicited invariably and had latencies > 2 ms. Monosynaptic i.p.s.ps seem not to fail and have latencies shorter than 1 ms (ref. 12).

Inhibitory coupling was assessed by averaging postsynaptic responses to bursts of presynaptic action potentials before and 12–20 min after tetanic stimulation. Recurrent inhibition was

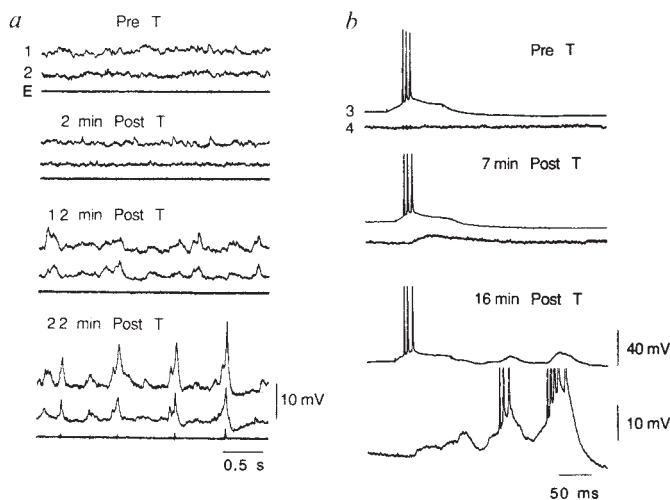


Fig. 4 Changes in activity of CA3 cells after tetani. *a*, Intracellular records from CA3 cells 1 and 2 and local extracellular field (E) in a longitudinal slice before, and at various times after, tetanic stimulation (50 Hz, 5 s, $\times 2$). Simultaneous e.p.s.ps were not detected before stimulation. The frequency of synaptic events was not increased at 2 min after tetanus. At 12 min some e.p.s.ps occurred simultaneously (within 5 ms) in both cells. The amplitude of simultaneous events was increased to 5–18 mV at 22 min after stimulation. Extracellular potentials were associated with the larger simultaneous events. Field potential amplitude and duration suggested the degree of synchrony in the CA3 cell population was much less than that during epileptiform discharges on pharmacological disinhibition. Both cells were hyperpolarized between -75 and -85 mV to prevent firing. *b*, Changes in the influence of cell 3 on cell 4 after mossy fibre tetani (20 Hz, 10 s, $\times 2$) in a transverse slice. Before stimulation cell 3 had no effect. At 7–9 min, firing in cell 3 evoked polysynaptic e.p.s.ps in 28 out of 54 trials at 0.5 Hz. At 15–17 min, cell 3 could elicit (in 8 of 60 trials) a barrage of synaptic events lasting about 200 ms which led to firing in cell 4 (action potentials cut). E.p.s.ps in cell 3 were synchronous with later events in cell 4.

reduced by $>20\%$ in 15 of 20 pairs and was unchanged in 5 cases (Fig. 3c). The mean amplitude of the averaged i.p.s.p. was 0.7 ± 0.3 mV before stimulation and 0.3 ± 0.2 mV after stimulation ($P < 0.02$, paired Mann-Whitney U test, $n = 20$). The input resistance of postsynaptic cells was not reduced (before stimulation, 29 ± 5 M Ω ; after, 31 ± 6 M Ω , $n = 20$). In five cells tested, no consistent change in reversal potential was detected: it was reduced in two cells, unchanged in two cells and increased in one cell with mean values of -72 ± 4 mV before, and -70 ± 3 mV after, tetani.

Although recurrent inhibition was suppressed, the site of modification in the inhibitory circuit is not known. Pyramidal-cell e.p.s.ps might evoke inhibitory cell firing less reliably²³ or a change might occur at inhibitory synapses^{17,18,20}. The inhibitory depression alone causes an increase in functional excitatory connectivity¹⁵. Transmission in polysynaptic pathways could also be enhanced by a potentiation of monosynaptic excitatory connections. A potentiated underlying e.p.s.p. might appear to reduce inhibitory coupling. This is unlikely to be generally true because usually e.p.s.ps were not seen at inhibitory connections when transmission failed after tetani (Fig. 3b).

What consequences arise from this plasticity in local synaptic circuits? Because activity can spread through poly-synaptic pathways (Fig. 1) the synchrony of spontaneous firing in the CA3 cell population should increase. In transverse and longitudinal slices simultaneous e.p.s.ps (within 5 ms) emerged in recordings from pairs of CA3 cells several minutes after tetanic stimuli (Fig. 4a). They were usually rhythmic with period 0.5–2 s and often grew in amplitude with time. In 14 of 24 cell pairs in transverse slices, and in 7 of 11 cell pairs in longitudinal slices, the amplitude of simultaneous e.p.s.ps exceeded 5 mV in both recorded cells. These events were larger than summed unitary e.p.s.ps (ref. 14) generated by a presynaptic burst at monosynaptic recurrent connections (up to 3 mV) and so must reflect simultaneous activity in several presynaptic neurons.

Several factors suggest that the simultaneously firing cells were CA3 cells. Simultaneous e.p.s.ps emerged after stimulating both mossy fibres and longitudinal fibres and so were not specific to the tetanized pathway. They were correlated with synchronous firing in some nearby cells because small field potentials were recorded extracellularly (Fig. 4a). Finally simultaneous e.p.s.ps could be initiated by activity induced in a single CA3 cell (Fig. 4b).

We have shown that latent excitatory pathways between burst firing CA3 cells may be uncovered after afferent tetanization. This plasticity in local synaptic circuits enhances the synchrony of firing in the CA3 cell population^{15,24}. Partly synchronous

firing of CA3 cells may also underlie the sharp waves recorded from the hippocampus of intact animals²⁵. Sharp waves are enhanced in amplitude and frequency after tetanic afferent stimulation *in vivo*²⁵.

Finally, it has been suggested that groups of synaptically associated cortical cells might represent a basis for information storage and might be formed if excitatory collateral synapses between them were strengthened^{6–10}. Then, subsequent excitation of some neurons of a group could spread by local excitatory pathways to initiate, or recall, activity in the other associated cells^{7–9}. We have shown that much of the neuronal machinery assumed in these predictions is present in the hippocampus. Divergent, local excitatory synapses exist^{13–15}. They may be persistently enhanced after afferent stimuli, at least in part, due to a depression of recurrent inhibition. When they are enhanced, groups of cells become associated and fire synchronously. Finally, activity elicited in a few cells²⁶ or even one cell^{15,27} can spread through newly functional excitatory synaptic pathways to initiate fully or partly synchronous firing (Fig. 4b) in a larger group of cells from the CA3 neuronal population.

We thank Roger D. Traub for discussion and encouragement. This work was supported by grants from the NIH and the Klingenstein Foundation.

Received 28 April; accepted 4 September 1987.

1. Tsukahara, N. *A. Rev. Neurosci.* **4**, 351–379 (1981).
2. Bliss, T. V. & Lomo, T. *J. J. Physiol., Lond.* **232**, 331–356 (1973).
3. Schwartzkroin, P. A. & Wester, K. *Brain Res.* **89**, 107–119 (1975).
4. Wigstrom, H. & Gustafsson, B. *Nature* **301**, 603–604 (1983).
5. Kelso, S. R., Ganong, A. & Brown, T. H. *Proc. natn. Acad. Sci. U.S.A.* **232**, 85–87 (1986).
6. Lorente de No, R. in *Physiology of the Nervous System* (ed. Fulton, J. F.) 288–312 (Oxford University Press, 1949).
7. Hebb, D. O. *The Organisation of Behaviour* (Wiley, New York, 1949).
8. Marr, D. *Phil. Trans. R. Soc. Lond.* **262**, 24–81 (1971).
9. Gardner-Medwin, A. R. *Proc. R. Soc. B* **194**, 375–402 (1976).
10. Abeles, M. *Local Cortical Circuits* (Springer, Berlin, 1982).
11. Andersen, P., Eccles, J. C. & Loynning, Y. *Nature* **198**, 540–542 (1963).
12. Miles, R. & Wong, R. K. S. *J. Physiol., Lond.* **356**, 97–113 (1984).
13. MacVicar, B. A. & Dudek, F. E. *Brain Res.* **184**, 220–223 (1980).
14. Miles, R. & Wong, R. K. S. *J. Physiol., Lond.* **373**, 397–418 (1986).
15. Miles, R. & Wong, R. K. S. *J. Physiol., Lond.* **388**, 611–630 (1987).
16. Getting, P. A. & Dekin, M. S. *J. Neurophysiol.* **53**, 466–480 (1985).
17. Yamamoto, C. & Chujo, T. *Expl. Neurol.* **58**, 242–250 (1978).
18. Misgeld, U., Sarvey, J. M. & Klee, M. R. *Expl. Brain Res.* **37**, 217–229 (1979).
19. Larsen, J. & Lynch, G. *Science* **232**, 985–988 (1986).
20. Steltzer, A., Slater, N. T. & ten Bruggencate, G. *Nature* **326**, 698–700 (1987).
21. Lorente de No, R. *J. Psychol. Neurol.* **46**, 113–117 (1934).
22. Bartesaghi, R., Gessi, T. & Sperti, L. *Expl. Neurol.* **82**, 550–567 (1983).
23. Zieglansberger, W., French, E. D., Siggins, G. R. & Bloom, F. E. *Science* **205**, 415–417 (1979).
24. Traub, R. D., Miles, R., Wong, R. K. S., Schulman, L. S. & Schneiderman, J. H. *J. Neurophysiol.* (in the press).
25. Buzsaki, G. *Brain Res.* **300**, 179–182 (1984).
26. Wong, R. K. S. & Traub, R. D. *J. Neurophysiol.* **49**, 442–458 (1983).
27. Miles, R. & Wong, R. K. S. *Nature* **306**, 371–373 (1983).

Corticofugal feedback influences the generation of length tuning in the visual pathway

P. C. Murphy & A. M. Sillito

Department of Visual Science, Institute of Ophthalmology, Judd Street, London WC1 9QS, UK

In the feline visual system, neurons exhibiting sensitivity to the length of a moving contour were first observed in the cortex^{1,2} and described as 'hypercomplex cells'. In these cells an increase in stimulus length beyond an optimal value leads to a rapid decline in response. This decline has been attributed to an intracortical inhibitory input¹ which may be driven by layer VI cells with very long receptive fields³. It is now clear, however, that cells in the dorsal lateral geniculate nucleus (dLGN), exhibit a degree of length tuning similar to that of cortical 'hypercomplex cells'^{4,5}, suggesting that this response property could be generated subcortically. Alternatively, as the dLGN receives a massive corticofugal projection from layer VI cells in the visual cortex⁶, it is possible that this input has a function in generating length tuning in the dLGN. We have investigated this issue by comparing the length tuning of dLGN cells with and without corticofugal feedback. The data show that corticofugal feedback makes a highly significant contribution to the length tuning of dLGN cells. This raises the possibility that length tuning is an emergent property of the geniculo-cortical loop.

Our report is based on length tuning curves obtained for 115 neurons recorded in eight anaesthetized cats, prepared using standard procedures for visual neurophysiology⁷. Glass micropipettes were used to make simultaneous recordings from cells in the A-laminae of the right and left dLGNs. Cells were classified as X- or Y-type using standard criteria^{8,9}. Receptive fields were routinely mapped with small flashing spots of light and checked for possible asymmetries in the field organization. The strength of the surround inhibition of centre responses was documented by plotting the response of cells to flashed spots of light of varying diameter. The fields were then studied with moving bars of light of varying length. When a series of dLGN cells on both sides of the brain had been studied, areas 17 and 18 of the visual cortex on one side were removed by aspiration. Further length tuning curves were plotted for dLGN cells on both sides. In some cases we could hold stable recordings of the same pair of cells before and after cortical aspiration. All recordings are from cells with receptive field eccentricities within 12° of the area centralis; histological analysis confirmed that the lesions extended well beyond those regions of areas 17 and 18 that would have been activated by our stimuli.

Our observations are summarized in Fig. 1. This compares the responses of two on-centre Y cells recorded in the same experiment. The responses of the cell with corticofugal feedback are totally suppressed at bar lengths of 2° and above, and those of the cell lacking feedback are reduced by up to 40% at bar lengths of 8° and above. Note that although one cell shows a brisk response to a 6° bar and the other no response, both show a similar degree of response attenuation to the 6° flashing spot. This suggests that loss of the corticofugal input has a much greater effect on the mechanisms generating the attenuation of responses to long bars than it does on the centre-surround antagonism revealed by stationary flashing spots of comparable diameter. As layer VI cells in the visual cortex are well activated by long moving bars but are poorly, if at all, driven by large flashing spots^{10,12}, this is not altogether surprising. Work in progress suggests that even in cells with intact corticofugal feedback, the degree of length tuning revealed by moving bar stimuli does not correlate in any simple way with the strength of centre-surround antagonism tested by flashing stimuli.

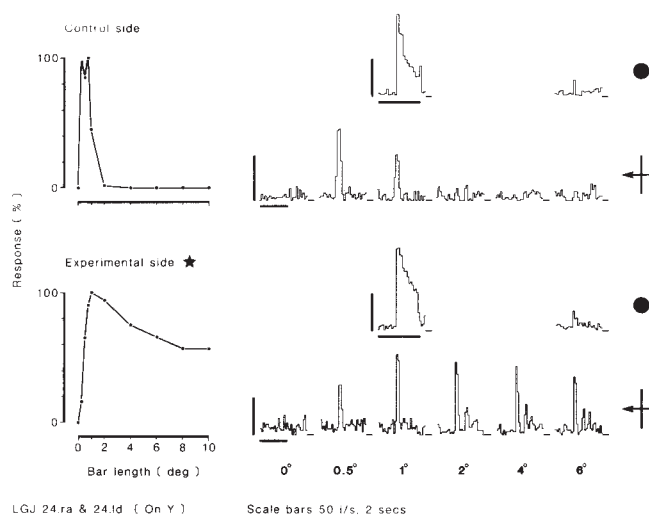


Fig. 1 Contribution of corticofugal feedback to the length tuning of dLGN cells. Data refers to two on-centre Y cells recorded from the left and right dorsal lateral geniculate nuclei of a single cat; one with (control side) and one without (experimental side) feedback from areas 17 and 18 of the visual cortex. Visual stimuli were presented on a modified Joyce display (Joyce Electronics, Cambridge) under computer control. Length-tuning curves were constructed for either a dark or light bar of varying length moving over the receptive field. To maintain comparability a constant stimulus velocity of 3.3° s^{-1} and contrast of either 150 on 50 cd m^{-2} or 50 on 150 cd m^{-2} (according to centre type) was used for all cells. The bar widths were either 0.3° or 0.5° (0.5° for the cells in this case) and the lengths were varied over $0\text{--}10^\circ$, with steps of 0.25° at the shorter lengths and 2° thereafter. The bar lengths were presented in an interleaved randomized sequence to minimize the effects of response variability. Responses were averaged over 5–20 trials (5 for this case), and assessed from the accumulated count in the bins constituting the response area in the peristimulus time histograms (PSTHs) after subtraction of the background discharge level. Length-tuning curves for the cells are illustrated on the left, with responses expressed as a percentage of the peak response. The cell on the control side is strongly end-stopped, but that on the experimental side is very much less so (see text). On the right are shown peristimulus time histograms (PSTHs) illustrating the responses to the moving bar from which the curves are constructed at six representative lengths (bar and arrow symbol), and the responses to spots of two different diameters flashed over the receptive fields and averaged for 20 trials. Scale bars: 50 impulses per second, 2 s. Note that both cells show a similar attenuation of the response to a 6° (centre plus surround) as compared with that to a 1° (centre only) spot. In all figures, the star indicates that the records refer to a cell without corticofugal feedback.

Recordings were obtained from three pairs of cells while the corticofugal input was removed. One pair (Fig. 2) were both on-centre X cells and under normal conditions exhibited about 80% attenuation of their peak responses with bars longer than 1° . Following removal of the corticofugal influence on the right side, the length tuning of the cell in the right dLGN was considerably reduced and, if anything, that of the cell in the left dLGN increased. These changes were not associated with any obvious shift in excitability or optimal response magnitude.

To quantify the data for the population of cells studied with and without corticofugal feedback we subdivided the cells into 10 categories, with those in category 10 having the strongest length preference and those in category 1 the weakest (Fig. 3). There is a significant ($p < 0.001$ student 't' test) shift of the distribution of length tuning in the population of cells without corticofugal feedback towards the lower values of the index. The average plateau response for the 56 cells studied without corticofugal feedback was 43% ($\pm 2\%$ s.e.m.) less than the peak, whereas that for the 59 cells with corticofugal feedback

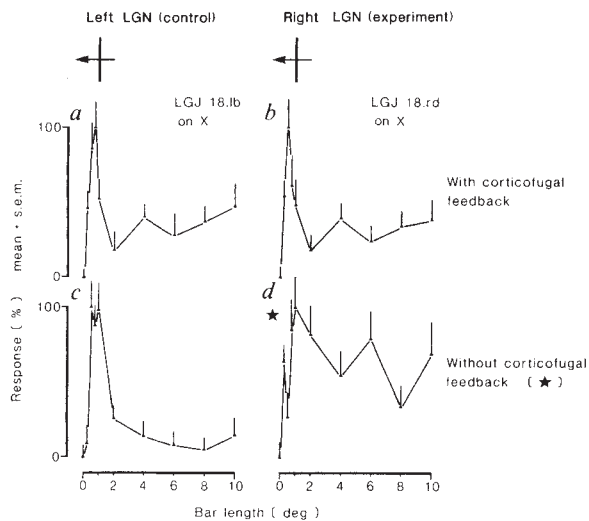


Fig. 2 Length-tuning curves for two on-centre X cells, recorded simultaneously from left and right dLGNs. Results were obtained from one of three attempts in which recordings of the pair of cells were successfully held while the corticofugal input was removed. The curves were constructed as in Fig. 1, except that the width of the stimulus was 0.3° in each of these cases. Error bars show standard error of the mean. Upper records (a and b): control situation, in which both cells were strongly end-stopped (see text). Lower records (c and d): tuning of the same two cells following removal of the cortical input to one side. On the experimental side (d) end-inhibition was roughly halved over the 2° , 4° and 6° range, without any change in the maximal response frequency, whereas on the side with intact corticofugal feedback (c) it was slightly increased.

was 70.5% ($\pm 2\%$ s.e.m.). We have also calculated the plateau responses from a range of alternative criteria, including the averaged values at 8° and 10° , and all show a significant difference between the two populations of cells. There was no significant difference between the average spontaneous activity or optimal response frequency in either group. We further checked that the shape of the tuning curves obtained in the absence of corticofugal feedback were not due to the use of saturating contrasts by ascertaining that higher contrasts produced larger responses and that the shape of the tuning profile also held for lower contrast stimuli.

The data for dLGN cells with corticofugal feedback reinforce and extend the observations of Cleland *et al.*⁴ regarding length tuning of dLGN cells. Cortical cells are classified as hypercomplex when the plateau response of the length-tuning curve is less than 50% of the peak, so a significant proportion of dLGN cells fall into the hypercomplex category on this criterion, as indicated by the shaded portions in Fig. 3. We suggest the length tuning exhibited by dLGN cells in the absence of corticofugal feedback follows simply from the various mechanisms known to underlie the concentric receptive field organization. This will involve two types of contribution from centre-surround antagonism, one deriving from interactions in the retina and one from the mechanisms in the dLGN (ref. 4). A narrow bar of light moving over the receptive field will activate a relatively smaller proportion of the surround mechanisms than the centre. Thus the component of surround antagonism of centre response will be much less than that seen with the large spot covering the entire surround. Our data suggest that the corticofugal feedback serves to redress the balance of the surround antagonism for bars, so that the receptive field selectivity to the dimension or position of an optimal stimulus is as good for moving contours as it is for flashing spots (Fig. 4). The corticofugal feedback appears to facilitate inhibitory mechanisms in the dLGN and generate an enhanced inhibitory drive, the tuning profile of which presumably reflects that of the layer VI cells providing

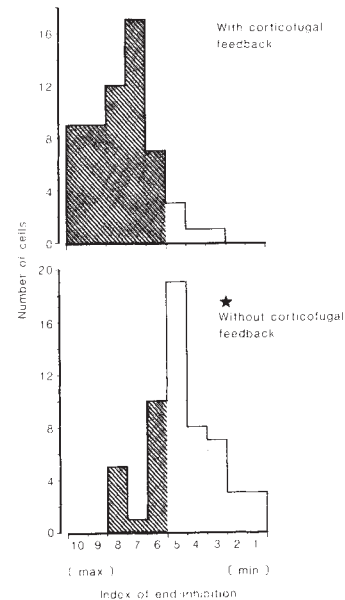


Fig. 3 Block histograms showing the distribution of length tuning in dLGN cells with and without corticofugal feedback. For comparison in this study, end-inhibition was quantified by calculating the difference between the peak response and the average response for 2° , 4° and 6° , expressed as a percentage of the peak: $(\text{peak} - \text{plateau}) \times 100 / \text{peak}$, where the plateau was the average response at 2° , 4° and 6° . This measure is sensitive to both observed effects of cortical ablation: an increase in the plateau response level, and a decrease in the tightness of the tuning peak. The cells were then subdivided into 10 categories of varying end-inhibition: category 1 represents cells with little or no end-stopping, and category 10 those showing an almost total suppression of response when bar lengths exceed 2° . Removing the cortical input produced a clear reduction in the degree of end-stopping, with the mean shifting from 71% ($\pm 2\%$ s.e.m., $N = 59$) to 43% ($\pm 2\%$ s.e.m., $N = 56$). Note that cells with greater than 50% end-inhibition, which would be considered hypercomplex in the visual cortex, are indicated by the shaded portions of both graphs.

the feedback. Thus it will be only weakly activated by flashing spots or annuli but optimally stimulated by moving bars or contours that provoke the best response from layer VI cortical cells. Although this could imply that the feedback should be orientation-tuned, the evidence does not favour this view. The corticofugal projection is retinotopically organized, with the representation of each retinal location in the cortex believed to contain a complete subset of orientation columns^{1,11}, and so the feedback probably derives from columns representing all orientations. Previous work on the corticofugal feedback to the dLGN in the cat¹² has shown that for an area of cortex representing 2 in visual space centred over a given retinotopic point in the dLGN, the feedback has an excitatory and an inhibitory influence, but beyond 2 the effect is inhibitory. Our present protocol failed to detect marked facilitatory effects but the data supports the inhibitory effect beyond 2° . Other studies have identified a range of variable effects on dLGN cell responses^{13,14}, but it is notable that the effect we find seems to be relatively consistent across the population of dLGN neurons, suggesting that it has detected some general feature of the corticofugal influence with respect to the response to moving bars.

Cleland *et al.*⁴ suggest that length tuning in the cortex comes from that present in the dLGN. Our data affect their argument by demonstrating that cortical mechanisms partially underlie the properties seen in the dLGN. A recent report³ found that localized blockade of the activity of layer VI cells eliminated the length tuning of cortical hypercomplex cells, which was explained in terms of a loss of a drive from layer VI cells to

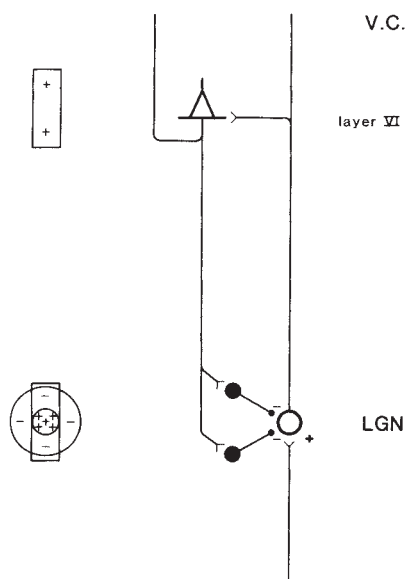


Fig. 4 Concept diagram summarizing the postulated influence of the corticofugal input on length tuning in the dLGN. To the right is shown a dLGN relay cell receiving excitatory input from a retinal afferent and in turn contacting a layer VI pyramidal cell in the visual cortex. The dLGN cell has a centre-surround type receptive field, as indicated to the left, whereas the cortical cell responds preferentially to elongated bars. The dLGN cell also receives an input from intrinsic inhibitory interneurons, which serve to enhance the surround suppression of centre response. Hence a flashing spot covering the entire receptive field will generate inhibition from the surround that almost cancels out the excitation from the centre (quantified here by the distribution of + and - signs in the centre and surround). A bar will cover less of the receptive field surround and so fail to engage the inhibitory mechanisms at retinal and geniculate level as effectively. However, it will also stimulate the cortical cell, generating an excitatory feedback to the dLGN which is presumed to enhance the effectiveness of the inhibitory input generated by the relay cell surround. Hence the cell will show similar degrees of response attenuation for bars and spots that extend beyond the receptive field centre.

inhibitory interneurons in layer IV of the cortex. Our observations suggest a similar function for the layer VI projection to the dLGN.

This raises questions about the contribution of dLGN-cell length tuning to that of cortical hypercomplex cells. It would be surprising if the length-tuning characterizing dLGN cell responses had nothing to do with the properties of end-stopped layer IV simple cells and indeed evidence supports a length-tuned excitatory input to cortical hypercomplex cells¹⁵. However, many cortical cells lack significant length tuning, and thus it seems necessary that a component of the synaptic organization in the cortex is arranged in a fashion which enables it to eliminate the bias present in the dLGN input. Possibly the length tuning conferred by the input from the dLGN to the cortex is entirely lost and reconstituted for some cells. The presence of length tuning in the dLGN and layer IV of the cortex questions the significance attached to this response property and the position of hypercomplex cells in the hierarchy of cortical processing. We consider that our observations demonstrate a cortical influence on dLGN cells which markedly shifts their response properties in relation to elongated visual contours moving over their receptive field, in a direction which appears to optimize the bias of the cell towards the spatial domain of the receptive field centre. As the output from the cortex in turn depends on the input from the dLGN, it follows that this response property emerges from the operation of a neuronal circuit, not from a hierarchical sequence.

We thank Miss H. E. Jones and acknowledge the MRC and Wellcome Trust for support.

Received 27 June; accepted 7 September 1987.

- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 229-289 (1965).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **195**, 215-243 (1968).
- Bolz, J. & Gilbert, C. D. *Nature* **320**, 362-365 (1986).
- Cleland, B. G., Lee, B. B. & Vidyasagar, T. R. *J. Neurosci.* **3**, 108-116 (1983).
- Jones, H. C. & Sillito, A. M. *J. Physiol., Lond.* **390**, 32P (1987).
- Gilbert, C. D. & Kelly, J. P. *J. comp. Neurol.* **163**, 81-106 (1975).
- Sillito, A. M. & Kemp, J. A. *Brain Res.* **277**, 63-77 (1983).
- Enroth-Cugell, C. & Robson, J. G. *J. Physiol., Lond.* **197**, 517-552 (1966).
- So, Y. T. & Shapley, R. M. *Expl Brain Res.* **36**, 533-550 (1979).
- Gilbert, C. D. *J. Physiol., Lond.* **268**, 391-421 (1977).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106-154 (1962).
- Tsumoto, T., Creutzfeldt, O. D. & Legendy, C. R. *Expl Brain Res.* **32**, 345-364 (1978).
- Geisert, E. E., Langsetmo, A. & Spear, P. D. *Brain Res.* **208**, 409-415 (1981).
- McClurkin, J. W. & Marrocco, R. T. *J. Physiol., Lond.* **348**, 135-152 (1984).
- Sillito, A. M. & Versiani, V. J. *J. Physiol., Lond.* **273**, 775-790 (1977).

Different *H-2* subregions influence immunization against retrovirus and immunosuppression

Richard P. Morrison*, Patricia L. Earl†, Jane Nishio*, Donald L. Lodmell*, Bernard Moss† & Bruce Chesebro*

* Laboratory of Persistent Viral Diseases, National Institutes of Health, Rocky Mountain Laboratories, Hamilton, Montana 59840, USA

† Laboratory of Viral Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Friend murine leukaemia virus complex (FV) causes an immunosuppressive retrovirus-induced disease. In certain mouse strains, FV shows striking similarities to human immunodeficiency virus (HIV) infection in man in that infected mice have severe T-cell immunosuppression but also develop virus-neutralizing antibodies incapable of eliminating infected cells. Previously we noted the influence of mouse major histocompatibility complex (*H-2*) genes on both FV-induced immunosuppression¹ and on ability to protect mice against FV by immunizing with a vaccinia-Friend murine leukaemia helper virus (F-MuLV) envelope (*env*) recombinant virus². Here we show that different subregions of *H-2* are involved in susceptibility to virus-induced immunosuppression (*H-2D* subregion) and protective immunization with a recombinant vaccinia virus (*H-2K* or *I-A* subregions). Thus, susceptibility to virus-induced immunosuppression does not preclude protection by vaccinia-Friend immunization. The mechanism of protection seems to involve priming of immune T cells, and not initial induction of neutralizing antibodies or cytotoxic T lymphocytes (CTL) (ref. 2). Subsequent virus challenge generates a secondary response, resulting in appearance of IgG antibodies and CTL. In human HIV infection there could also be host genetic influences on elements of disease pathogenesis, such as immunosuppression, and on the success of T-cell priming by potential protective vaccines.

FV causes erythroleukaemia in susceptible strains of mice. The disease occurs after infection of immunocompetent adult mice, usually resulting in death within one to three months³. The virus induces immunosuppression to nonviral antigens. T-cell helper function is decreased¹, but certain mouse strains can develop an antiviral humoral immune response starting 15-20 days after infection¹. This response consists of virus-neutralizing antibodies and cytotoxic antibodies capable of lysing virus-infected cells in the presence of complement⁴⁻⁷. Appearance of these antibodies is controlled by a single gene, *Rfv-3*, and correlates kinetically with the clearing of viraemia⁷. Despite the presence of antiviral antibodies leukaemic splenomegaly persists, but infected leukaemic cells no longer display virus-encoded cell surface antigens and release little infectious virus^{8,9}. This paradoxical antiviral immune response in the presence of virus-induced immunosuppression to nonviral antigens is remarkably similar to the pattern observed in many

Table 1 Protective immunization against herpes simplex virus (HSV-2) using a vaccinia recombinant

Immunization	Number dead/total	
	<i>H</i> -2 ^{a/b}	<i>H</i> -2 ^{a/a}
Vaccinia-HSV <i>env</i>	4/20*	1/20*
Vaccinia-F-MuLV <i>env</i>	20/20	20/20
None	23/23	8/8

Groups of male and female (B10.A × A.BY)_F₁ (*H*-2^{a/b}), and (B10.A × A/WySn)_F₁ (*H*-2^{a/a}) mice were immunized by tail scratch scarification with either live recombinant vaccinia-HSV *env* (ref. 16) or vaccinia-F-MuLV *env* virus as described in Fig. 1 legend; 24 d after immunization mice were challenged intraperitoneally with 5 × 10⁷ plaque-forming units (PFU) of HSV-2. Results are expressed as the number of dead animals divided by the total at 21 d after HSV-2 challenge.

* $p < 0.05$ by χ^2 test for a 2 × 2 table.

human AIDS (acquired immune deficiency syndrome) patients¹⁰⁻¹⁵. Furthermore, the inability of virus-specific neutralizing or cytotoxic antibodies to eliminate infected cells is also seen in HIV infection in man. In both FV and HIV infections, it is likely that cell-mediated immune mechanisms are required to eliminate infected cells. For these reasons the FV system could provide a model for development of a vaccine in an immunosuppressive situation.

Previous studies on FV-induced immunosuppression showed that the immunoglobulin G (IgG) anti-sheep red blood cell (SRBC) response was suppressed after FV inoculation in *H*-2^{a/a} mice, and that this response was not suppressed in *H*-2^{a/b} mice¹. Protection following tail-scratch inoculation of a vaccinia-Friend murine leukaemia helper virus (F-MuLV) envelope (*env*) recombinant virus was also influenced by genes of the major histocompatibility complex (*H*-2) (ref. 2). Mice having the *H*-2^{a/b} genotype were protected by immunization with the recombinant virus, but *H*-2^{a/a} mice were not (Fig. 1 and ref. 2). Thus, susceptibility to FV-induced immunosuppression was apparently associated with an inability to induce protection with the vaccinia-F-MuLV *env* recombinant vaccine. To investigate this possibility, we used congenic *H*-2 recombinant mice to determine the region(s) of *H*-2 involved in both protective immunization and FV-induced immunosuppression. Protection with the vaccinia-F-MuLV *env* recombinant was observed in all strains of mice having the heterozygous *H*-2^{a/b} genotype at the *K* and *I*-A regions. Mice having the homozygous *H*-2K, *I*-A^{a/a} genotype were not protected (Fig. 1). In a second experiment, FV-infected and normal unimmunized *H*-2 congenic

Table 3 Protective immunization against challenge with live allogeneic FV-releasing leukaemia cells

Immunization	Number leukaemic/total	
	<i>H</i> -2 ^{a/b}	<i>H</i> -2 ^{a/a}
FV-N	0/18*	0/17*
Fixed F-MuLV in CFA	0/19*	1/20*
Vaccinia-F-MuLV <i>env</i>	0/20*	19/20
Vaccinia-influenza HA	18/20	17/19
None	17/20	20/20

Congenic (B10.A × A.BY)_F₁ and (B10.A × A/WySn)_F₁ mice were immunized with either FV-N, formalin fixed F-MuLV in CFA or vaccinia-F-MuLV *env* as described in Fig. 1. Mice were immunized by exposing a needle scratch at the base of the tail to 1 μ l vaccinia-influenza HA containing 10⁷ PFU. 21 d after immunization, mice were challenged intravenously with 2.5 × 10⁷ allogeneic FV-releasing leukaemia cells (BALB.K-fv-1, *H*-2^k) (ref. 23). Data are presented as the number of leukaemic or dead mice at 150 d after challenge divided by the total.

* Values are significantly different from controls as determined by the χ^2 test for a 2 × 2 table ($p < 0.001$).

mouse strains were tested for FV-induced suppression of the anti-SRBC response (Fig. 1). In contrast to the results on protection by vaccinia-F-MuLV *env*, susceptibility to FV-induced immunosuppression was influenced by the *H*-2D region. Homozygous *H*-2D^{a/a} mice had significantly suppressed anti-SRBC responses, whereas *H*-2D^{a/b} mice were not immunosuppressed. In the case of (B10.A(5R) × A/WySn)_F₁ and (B10.A(18R) × A/WySn)_F₁ mice, susceptibility to FV-induced immunosuppression (*H*-2D^{a/a} genotype) did not preclude protective immunization with the vaccinia-F-MuLV *env* recombinant (*H*-2K and *I*-A^{a/b} genotype) (Fig. 1).

To determine whether the lack of protection in *H*-2K, *I*-A^{a/a} mice was due to an inability to be immunized with any vaccinia recombinant viruses, we attempted to protect *H*-2^{a/a} mice against herpes simplex virus type-2 (HSV-2) infection by immunizing with a vaccinia-HSV *env* recombinant¹⁶. The effect associated with the *H*-2 gene found with the vaccinia-F-MuLV *env* recombinant was not observed with the vaccinia-HSV *env* recombinant (Table 1). Therefore, the *H*-2 difference observed with vaccinia-F-MuLV *env* appeared to involve the immune response to the F-MuLV *env* protein rather than the response to recombinant vaccinia virus.

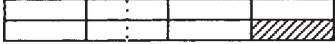
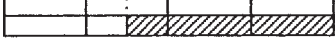
The influence of the *H*-2K or *I*-A regions on immunological priming by vaccinia-F-MuLV *env* suggested that there was a strong immune response (*I*r) gene effect with this vaccine. As

Table 2 Effect of immunization with vaccinia-F-MuLV *env* on immunoglobulin class and titre of Friend virus neutralizing antibodies

<i>H</i> -2K, <i>I</i> -A genotype	Vaccinia-influenza HA						Vaccinia-F-MuLV <i>env</i>					
	0 d		10 d		28 d		0 d		10 d		28 d	
	IgM	IgG	IgM	IgG	IgM	IgG	IgM	IgG	IgM	IgG	IgM	IgG
a/b	<2	<2	<2	<2	531*	<2	<2	<2	2	42*	2	128*
a/a	<2	<2	<2	<2	486*	<2	<2	<2	<2	<2	256*	3

The β -mercaptoethanol (ME) sensitivity of F-MuLV neutralizing antibody in vaccinia-F-MuLV *env* recombinant- and vaccinia-influenza haemagglutinin (HA) (ref. 21) recombinant-immunized mice was assayed at 0, 10 and 28 d after FV challenge. Mice were immunized and challenged as in Fig. 1. Samples of plasma were diluted 1:1 with PBS made 0.2 M in ME, incubated at 37 °C for 30 min, then further diluted in PBS now 0.01 M in ME, and assayed for neutralizing antibody². This treatment had no effect on virus viability and completely inhibited F-MuLV neutralization by an IgM monoclonal antibody against F-MuLV *env* (mAb 307) and had no effect on neutralization by an IgG_{2a} monoclonal antibody against F-MuLV *env* (mAb 48) (ref. 22). The IgG anti-F-MuLV antibody titre represents the titre obtained from ME-treated plasma. The IgM anti-F-MuLV antibody titre was the difference between the ME-treated plasma titre and the titre obtained with untreated plasma. Mice were grouped as to their *H*-2K and *I*-A genotype: *H*-2K, *I*-A^{a/b} includes (B10.A × A.BY)_F₁, (B10.A(5R) × A/WySn)_F₁ and (B10.A(18R) × A/WySn)_F₁ mice and *H*-2K, *I*-A^{a/a} includes (B10.A × A/WySn)_F₁, (B10.A(1R) × A/WySn)_F₁, (B10.A(2R) × A/WySn)_F₁ and (B10.A(4R) × A/WySn)_F₁ mice. Groups consisted of 8-14 mice. Standard error values in all cases were less than one twofold dilution and are omitted from the table. Results for unimmunized mice were similar to results from mice immunized with vaccinia-influenza HA.

* Neutralizing antibody titres significantly ($P < 0.01$) greater than normal control mice by the Student *t*-test.

STRAIN	H-2 GENOTYPE	LEUKAEMIA INCIDENCE			PLASMA ANTI-SRBC TITRE	
		IMMUNIZATION			FV INFECTED	
		FV-N	FIXED F-MuLV IN CFA	VACCINIA- F-MuLV <i>env</i>		NORMAL
(B10.A X A.BY) _{F1}		1/66(2%)	0/30(0%)	4/67 (6%)	1825	2632
(B10.A X A/WySn) _{F1}		0/56(0%)	2/30(7%)	38/46(83%)	63	3286
(B10.A(1R) X A/WySn) _{F1}		0/10(0%)	0/10(0%)	23/31(74%)	1855	2897
(B10.A(2R) X A/WySn) _{F1}		0/10(0%)	2/10(20%)	23/30(77%)	2403	3490
(B10.A(4R) X A/WySn) _{F1}		0/10(0%)	0/9(0%)	19/27(70%)	2048	4096
(B10.A(5R) X A/WySn) _{F1}		1/9(11%)	0/9(0%)	4/30 (13%)	92	4872
(B10.A(18R) X A/WySn) _{F1}		1/10(10%)	0/9(0%)	6/30 (20%)	63	5070

 H-2^a
 H-2^b

Fig. 1 Identification of regions of the *H-2* complex involved in protective immunity and in FV-induced immunosuppression.

Methods. Congenic *F₁* hybrid mice were bred from B10.A, B10.A(2R), A.BY, A/WySn (Jackson, Maine), B10.A(1R), B10.A(4R), B10.A(5R) and B10.A(18R) (kindly provided by Jack Stimpfling). For the protection experiments, mice were immunized as follows: 1 µl containing 1×10^7 PFU of recombinant vaccinia-F-MuLV *env* virus² on to a 3-mm needle-scratch in the tail skin; 5×10^3 spleen focus forming units (SFFU) of live FV-N given intravenously; formalin fixed F-MuLV (44 µg) in Freund's complete adjuvant (CFA) injected into the rear footpads. 21 d after vaccination, mice were challenged intravenously with 2×10^4 SFFU of FV-B complex. All mouse strains tested are Fv-1^{b/b}, and thus FV-B complex replication is not restricted. Data are shown as the number of mice with leukaemic splenomegaly or dead divided by the total number of mice tested, at 150 d after FV challenge. Mice were examined weekly under ether anaesthesia by spleen palpation¹⁹. The percentage of dead or leukaemic mice is shown in parentheses. More than 95% of the unimmunized control mice and mice immunized with a control vaccinia recombinant (vaccinia-influenza-HA) (ref. 21) died of FV disease. FV-induced immunosuppression was examined by assaying the anti-SRBC plasma antibody titre 25 d after infection with FV (ref. 1). Immunosuppression data are expressed as the geometric means of the anti-SRBC plasma antibody titre of normal and FV-infected mice (10–16 mice per group). Protection values enclosed in boxes are not significantly protected ($p > 0.05$) compared to unimmunized mice (χ^2 test for a 2×2 table), whereas immunosuppression data enclosed in a box are significantly different ($p < 0.05$) from normal mice (Student *t*-test). All FV-uninoculated mice given vaccinia recombinant virus had significant amounts of anti-vaccinia neutralizing antibody and had normal anti-SRBC antibody responses 21 d after immunization with the vaccinia-F-MuLV *env* recombinant (data not shown). Therefore, failure to protect some mouse strains by immunizing with the vaccinia-F-MuLV *env* recombinant was not due to vaccinia virus-induced immunosuppression or to lack of immunological responsiveness to vaccinia virus.

poor immune responsiveness controlled by *Ir* genes can often be overcome by increasing the dose of immunogen, or by using immunogenic carrier proteins or more potent adjuvants^{17,18}, immunization was attempted with live attenuated FV (FV-N), and with formalin-inactivated F-MuLV in Freund's complete adjuvant. Protective immunity was demonstrated in all mouse strains, regardless of *H-2* genotype, when immunized with either agent (Fig. 1). Thus, the influence of *H-2K* or *I-A* genes on the ability to immunize mice with vaccinia-F-MuLV *env* was overcome.

We went on to determine whether this genetic effect on the protection after immunization the vaccinia-F-MuLV *env* recombinant was reflected in the humoral immune response to FV after immunization and challenge. The results indicated that early IgG FV-neutralizing antibodies were seen only in H-2K, I-A^{a/b} mice immunized with vaccinia-F-MuLV *env* (Table 2). Successful immunological priming was thus associated with an early IgG response, and with the early appearance of virus-specific CTL after challenge². Although protection could be mediated by either CTL or neutralizing antibodies, it is likely that CTL are the more critical effectors, because antibodies alone do not induce recovery in unvaccinated mice⁴⁻⁷.

As retroviral infections are thought to be primarily transmitted by allogeneic virus-infected cells, different immunization pro-

ocols were tested for efficiency in protecting mice against a lethal challenge of FV-infected leukaemia cells. The protection pattern was similar to that after challenge with cell-free virus (Table 3). These immunizations were therefore effective even against challenge with allogeneic virus-infected cells.

The data show that different *H-2* subregions are responsible for FV-induced immunosuppression and inability to prime certain mouse strains with vaccinia-F-MuLV *env*, and that failure of protective immunization by the recombinant vaccinia virus is not related to FV-induced immunosuppression. The *H-2D* gene affects spontaneous recovery from FV leukaemia by altering the kinetics of the virus-specific L3T4-positive T-cell response (refs 19, 20 and our unpublished results); following FV infection H-2D^{a/b} mice respond slowly and fail to recover from leukaemia²⁰. Possibly the delayed immune response in H-2D^{a/b} mice could produce lymphokines helpful in preventing immunosuppression.

Our results illustrate the complexity of genetically controlled host responses to retroviral infection and show that absence of some host-resistance genes can be overcome by vaccination using appropriate preparations. Similar effects might occur during HIV infection of humans and should be considered in the development of a vaccine for protection against AIDS.

We thank Robert Evans for graphic assistance, Irene Cook

Rodriguez for preparing the manuscript, and Drs John Portis and Masaaki Miyazawa for discussion.

Received 28 May; accepted 9 September 1987.

1. Morrison, R. P., Nishio, J. & Chesebro, B. *J. exp. Med.* **163**, 301-314 (1986).
2. Earl, P. L. *et al. Science* **234**, 728-731 (1986).
3. Teich, N., Wyke, J., Mak, T., Bernstein, A. & Hardy, W. in *RNA Tumor Viruses, Molecular Biology of Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 785-998 (Cold Spring Harbor Laboratory, New York, 1984).
4. Chesebro, B. & Wehrly, K. *J. exp. Med.* **143**, 73-84 (1976).
5. Chesebro, B., Bloom, M., Wehrly, K. & Nishio, J. *J. Virol.* **32**, 832-837 (1979).
6. Doig, D. & Chesebro, B. *J. exp. Med.* **150**, 10-19 (1979).
7. Chesebro, B. & Wehrly, K. *Proc. natn. Acad. Sci. U.S.A.* **76**, 425-429 (1979).
8. Doig, D. & Chesebro, B. *J. exp. Med.* **148**, 1109-1121 (1978).

9. Chesebro, B., Wehrly, K., Doig, D. & Nishio, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5784-5788 (1979).
10. Weiss, R. A. *et al. Nature* **316**, 69-72 (1985).
11. Robert-Guroff, M., Brown, M. & Gallo, R. C. *Nature* **316**, 72-74 (1985).
12. McDougal, J. S. *et al. J. Immunol. Meth.* **76**, 171-183 (1985).
13. Essex, M. *et al. Science* **220**, 859-862 (1983).
14. Lane, H. C. & Fauci, A. S. *A. Rev. Immun.* **3**, 477-500 (1985).
15. Rook, A. H. *et al. J. Immun.* **138**, 1064-1067 (1987).
16. Cremer, K. J., Mackett, M., Wohlenberg, C., Notkins, A. L. & Moss, B. *Science* **228**, 737-740 (1985).
17. Green, I., Paul, W. E. & Benacerraf, B. *J. exp. Med.* **123**, 859-879 (1966).
18. McDevitt, H. O. *J. Immun.* **100**, 485-492 (1968).
19. Chesebro, B., Wehrly, K. & Stimpfing, J. *J. exp. Med.* **140**, 1457-1467 (1974).
20. Britt, W. J. & Chesebro, B. *J. exp. Med.* **157**, 1736-1745 (1983).
21. Smith, G. L., Murphy, B. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7155-7159 (1983).
22. Chesebro, B., Britt, W., Evans, L., Wehrly, K., Nishio, J. & Cloyd, M. *Virology* **127**, 134-148 (1983).
23. Wolfe, J. H., Blank, K. J. & Pincus, T. *Immunogenetics* **12**, 187-190 (1981).

Two homologous protein components of hepatic gap junctions

B. Nicholson‡, R. Dermietzel*, D. Teplow, O. Traub†, K. Willecke† & J.-P. Revel

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
Institutes of * Anatomy and † Cell Biology, Universitätsklinikum, 4300 Essen 1, FRG

Gap junctions consist of closely packed pairs of transmembrane channels, the connexons, through which materials of low relative molecular mass diffuse from the cell to neighbouring cells. In liver, connexons consist of six protein subunits^{1,2} which, until now, were believed to be identical³. However, besides the major polypeptide of relative molecular mass (M_r) 28,000 (and see refs 4 and 6), a component of M_r 21,000 (21K) has been repeatedly observed³⁻⁵ in liver. The amino-terminal sequence (18 residues) of this less abundant protein shows that it is related to, but distinct from, the M_r 28K protein. Immuno-staining and immuno-precipitation show both proteins to be in the same gap junctional plaques. Thus, it seems that hepatic gap junction channels (and by extension possibly others) are composed of two (or more) homologous proteins.

In general, isolated hepatic gap junctions^{6,7} are found to contain a predominant protein of $M_r \sim 28K$, although an unrelated, M_r 16K, protein has been reported in protease resistant

fractions isolated by a specific protocol⁸. The 28K hepatic junction protein (HJP28) is related to the cardiac gap junction protein CJP47 (refs 9 and 10). The gene that encodes it has been cloned¹¹⁻¹³, and junction-like channels reconstituted¹⁴⁻¹⁶.

Purified fractions from rat or mouse liver typically contain degradation or aggregation products of HJP28 (refs 4 and 5). In addition there is a persistent 21K component (HJP21) which makes up <10% of the protein in rat but as much as 35% of it in mouse liver junctional fractions (Fig. 1a). Although antibodies raised against HJP21 do not react with HJP28 (ref. 17), or vice versa (Fig. 1b), a degree of homology is evident from peptide mapping^{4,5}.

Amounts of HJP28 and HJP21 sufficient to determine the sequence of 14-18 amino-terminal residues were obtained by electroelution from bands excised from SDS polyacrylamide gels¹⁸ or by direct Western blotting onto an activated glass fibre disk¹⁹. Comparison of the HJP21 sequence with that of HJP28 (Fig. 2) reveals that 9 of 18 residues are identical, and three more are very similar. Only minor differences are detected in comparisons between rat and mouse proteins (Fig. 3). The sequence proves that HJP21 cannot be derived from HJP28 by proteolysis. Comparisons with CJP47 from heart indicates that gap junction proteins form a divergent family.

The location of these proteins was studied using a rabbit polyclonal antibody against gel-isolated mouse HJP21 (ref. 17) and a rat monoclonal antibody against mouse HJP28 (refs 20 and 21). Both bind specifically to gap junctions *in situ* and to the appropriate protein on Western blots (Fig. 1b). Following affinity purification neither antibody cross-reacted with the heterologous protein in the denatured (Fig. 1b) or native state. Sections of mouse liver first treated with the monoclonal antibody against HJP28 and anti-HJP21 serum, followed by specific second antibodies, show a punctate pattern of fluorescent label-

Table 1 Immunoprecipitation of mouse liver gap junctions with monoclonal antibody against the 28K component

Experiment	Antibody	HJP28:HJP21 ratio*			Enrichment† for HJP28 (bound/unbound)
		Starting material	Unbound	Bound	
1	Anti-HJP28	1.3	1.2	2.5	2.1
	Non-specific IgG		1.3	ND	—
2	Anti-HJP28	1.5	1.3	2.9	2.2
	Non-specific IgG		1.2	1.1	0.9

In experiment 1, monoclonal antibodies against HJP28, or non-specific bovine immunoglobulin G (IgG), were bound to CNBr-activated Sepharose 4B. The derivatized beads were reacted with an excess of a mouse liver gap junction fraction and the HJP28:HJP21 ratio measured. After settling, the supernatant was analysed for unbound material. Following several washes of the beads in PBS, pH 7.4, 0.05% Tween 20, specifically bound junctions were eluted with 3 M KSCN for 5 min. The supernatant was collected and the elution repeated. An average from the analyses of both fractions is shown as the bound fraction. In experiment 2, a mouse gap junction fraction (pre-treated with 0.25% bovine serum albumin (BSA) in PBS to reduce non-specific adhesion of IgG to junctions) was reacted in suspension with either monoclonal antibodies against HJP28 or a non-specific IgG. Junctions were removed by centrifugation and reacted with immunobeads (Biorad) containing bound goat anti-rat IgG. Subsequent separations and elutions were as in experiment 1. ND, not detected.

* Ratios of HJP28 to HJP21 were determined by quantitative densitometry of silver-stained SDS-polyacrylamide gels.

† Enrichment for HJP28 was defined as the fractional increase in the HJP28:HJP21 ratio in the bound material compared to the unbound fraction. Unbound rather than untreated (or original) material was used as a standard because gap junction protein was in excess over antibody and this controlled for the non-specific and irreversible trapping of junctions by the solid support.

‡ Present address: Department of Biological Sciences, State University of New York at Buffalo, Buffalo, New York 14260, USA.

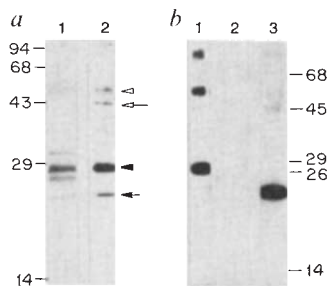


Fig. 1 *a*, Coomassie stained gels of lane 1, rat and lane 2, mouse liver gap junction fractions. HJP28 (filled arrowhead) and HJP21 (filled arrow) are present in 10/1 ratio in rat and 2/1 in mouse junctional fractions. Both proteins aggregate in SDS to form apparent homodimers marked with an open arrowhead and open arrow respectively. Lane 1 shows degradation products of HJP28 (M_r 26K–24K) and a minor amount of uricase (M_r 34K). M_r was determined by comparison with standards. *b*, Autoradiogram of Western blots of mouse liver gap junction fractions (similar to lane 2 in *a*, but at lower concentration) treated with lane 1, monoclonal anti-HJP28 antibody; lane 2, preimmune serum; and lane 3, anti-HJP21 serum. Size marked in *a* and *b* show M_r in thousands. **Methods.** Gap junction fractions were prepared according to established procedures^{4,6}. For rat isolates, the final sucrose gradient⁶ was substituted with a 35–49% (w/v) sucrose step gradient containing 0.3% deoxycholate. Conditions for Western blots²¹ and preparation of antisera^{17,21} have been described. Anti-HJP21 serum was absorbed out on an HJP28 affinity column to eliminate minor crossreactivity¹⁷.

ling (Fig. 3). Variations in the relative intensity of labelling of individual junctions can be detected, especially in rat liver, but it is clear that HJP21 and HJP28 colocalize precisely to the same gap junctional plaques.

To examine the relationship between the two proteins, the monoclonal antibody against HJP28 was bound to a solid support either directly or through secondary antibodies (Table 1), then used to precipitate junction plaques isolated from mouse liver. With either method the junctions bound to the beads were about two-fold enriched in HJP28, whereas those remaining in the supernatant contained relatively more HJP21. The differences in staining reported above may reflect the different ratios of HJP21 and HJP28 in individual junction plaques. Individual connexons could contain different numbers of 21K and 28K subunits; alternately gap junctions could contain two different types of connexons each made of only one type of subunit. Aggregation of junction protein in SDS leads to the formation of dimers (Fig. 1; ref. 4) in approximately the same ratio as the two monomers. No heterodimer is detected by Coomassie staining (Fig. 1*a*), although a small amount can occasionally be seen in Western blots (Fig. 1*b*). This suggests that the connexons may be predominantly homopolymers. It seems unlikely that

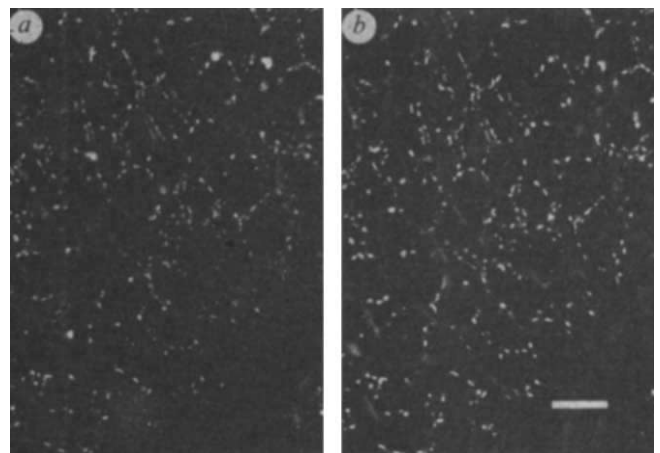


Fig. 3 *a*, *b*, Double immunofluorescent labelling of the same field of a mouse liver section stained with *a*, the monoclonal anti-HJP28 antibody and *b*, polyclonal anti-HJP21 antibody. The specific antibodies against the two proteins clearly stain the same structures throughout the section, although minor variations in the relative staining intensity can be seen. Scale bar, 10 μ m.

Methods. Cryosections were fixed in absolute ethanol for 15 min at 25 °C, rinsed in PBS and PBS with 0.1% BSA, and incubated with both monoclonal anti-HJP28 and polyclonal antibodies against HJP21. After additional washes, the sections were incubated with secondary antibodies (goat anti-rat IgG conjugated to rhodamine and goat anti-rabbit IgG conjugated to fluorescein). Controls, including using preimmune serum and tests for cross-species reactivity of the secondary antibodies, were negative.

HJP21 represents a membrane-bound accessory protein which is not part of the connexons, because of the high sequence similarity between the two hepatic junction proteins and other similarities^{4–7}.

Models for liver gap junctions^{1,2}, and possibly those for other junctions as well^{22,23}, will have to be made more complex. To determine the different functions and properties conferred on intercellular channels by the two homologous proteins will be of some interest^{24,25}.

We thank Dr R. Aebersold for help with blotting, B. Kunz and J. Pancok for technical assistance, and Dr B. Yancey and J. Hoh for helpful discussions. This work was supported by NIH grants to J.P.R.; by a grant from the American Heart Association, Los Angeles Affiliate (B.N.); and by the Deutsche Forschungsgemeinschaft (K.W. and R.D.), the Dr Mildred Scheel Stiftung-Deutsche Krebshilfe and the Fonds der Chemischen Industrie (K.W.).

Received 21 July; accepted 2 September 1987.

- Caspar, D. L. D., Goodenough, D. A., Makowski, L. & Phillips, W. C. *J. Cell Biol.* **74**, 605–628 (1977).
- Unwin, P. N. T. & Zampighi, G. *Nature* **283**, 545–549 (1980).
- Revel, J. P., Nicholson, B. & Yancey, S. B. *A. Rev. Physiol.* **47**, 263–279 (1985).
- Henderson, D., Eibl, H. & Weber, K. *J. molec. Biol.* **132**, 193–218 (1979).
- Nicholson, B., Hunkapiller, M., Grim, L., Hood, L. & Revel, J. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7594–7598 (1981).
- Hertzberg, E. L. *J. biol. Chem.* **259**, 9936–9943 (1984).
- Nicholson, B. J. & Revel, J. P. *Meth. Enzym.* **98**, 519–537 (1983).
- Finbow, M. E., Shuttleworth, J., Hamilton, A. E. & Pitts, J. D. *EMBO J.* **2**, 1479–1486 (1983).
- Nicholson, B., Gros, D., Kent, S., Hood, L. & Revel, J. P. *J. biol. Chem.* **260**, 6514–6517 (1985).
- Manjunath, C. K. *et al. Biochem. biophys. Res. Commun.* **142**, 228–234 (1987).
- Paul, D., *J. Cell Biol.* **103**, 123–134 (1986).
- Kumar, N. M. & Gilula, N. B. *J. Cell Biol.* **103**, 767–776 (1986).
- Heykes, R., Kozjek, G., Traub, O. & Willecke, K. *FEBS Lett.* **205**, 56–60 (1986).
- Spray, D. C., Saez, J. C., Brosius, D., Bennett, M. W. L. & Hertzberg, E. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5494–5497 (1986).
- Young, J. D., Cohn, Z. A. & Gilula, N. B. *Cell* **48**, 733–743 (1987).
- Dahl, G., Miller, T., Paul, D., Voellmy, R. & Werner, R. *Science* **236**, 1290–1293 (1987).
- Traub, O., Janssen-Timmen, U., Druege, P. M., Dermietzel, R. & Willecke, K. *J. cell. Biochem.* **19**, 27–44 (1982).
- Hunkapiller, M., Strickler, J. E. & Wilson, K. J. *Science* **227**, 304–311 (1984).

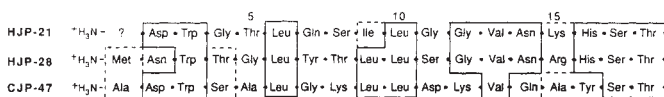


Fig. 2 The amino-terminal 18 residues of the three known gap junction proteins from (rat hepatic HJP21 and HJP28 and cardiac CJP47) share 30% identity overall (50% between the hepatic proteins). Solid boxes, identical residues; broken boxes, residues of similar character. High background typical of electroeluted samples¹⁸ made assignment of residue 1 for the rat HJP21 protein uncertain, but a methionine was identified in mouse HJP21. Alanine at position 16 in the rat protein was tentatively identified as histidine in the mouse. There are no differences between rat and mouse HJP28, except that the rat N-terminal methionine is lost from the mouse sequence, which begins with an asparagine (position 2 of the rat sequence).

19. Aebersold, R., Teplow, D., Hood, L. & Kent, S. *J. biol. Chem.* **261**, 4229-4238 (1986).
20. Willecke, K. *et al.* in *Gap Junctions* (eds Bennett, M. V. L. & Spray, D.) 67-76 (Cold Spring Harbor Laboratory, New York, 1985).
21. Janssen-Timmen, U., Traub, O., Dermietzel, R., Rabes, H. & Willecke, K. *Carcinogenesis* **7**, 1475-1482 (1986).
22. Gorin, M. B., Yancey, S. B., Cline, J., Revel, J. P. & Horwitz, J. *Cell* **39**, 49-58 (1984).
23. Revel, J. P. & Yancey, S. B. in *Gap Junctions* (eds Bennett, M. K. L. & Spray, D.) 33-48 (Cold Spring Harbor Laboratories, New York, 1985).
24. Traub, O., Druge, P. M. & Willecke, K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 755-759 (1983).

Oestrogen and glucocorticoid responsive elements are closely related but distinct

Gerd Klock, Uwe Strähle & Günther Schütz

Institute of Cell and Tumor Biology, German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

DNA sequences recognized by the glucocorticoid receptor¹⁻⁴ are termed glucocorticoid-responsive elements¹ because of their stimulatory effect on transcription. An oligonucleotide of 15 base pairs having partial or perfect symmetry is necessary for glucocorticoid induction⁵ and this same oligonucleotide is surprisingly also recognized by the progesterone receptor⁵. Here we define a palindromic sequence of 15 base pairs, modelled after a sequence element shared by the vitellogenin genes of frog and chicken⁶, which confers oestrogen inducibility on a heterologous promoter and can be converted into a glucocorticoid-responsive element by substitution of one or two bases at homologous positions in the palindrome. Considered with the observation that the DNA-binding domains of steroid receptors are closely related⁷, this finding demonstrates that the steroid-responsive elements constitute a family of related DNA sequences.

Recently we have demonstrated that a 15-base-pair (bp) oligomer with partial symmetry, and modelled upon the glucocorticoid-responsive element (GRE) of the rat tyrosine aminotransferase (TAT) gene, is sufficient to confer glucocorticoid hormone inducibility to a heterologous promoter⁵ (Fig. 1a). This oligonucleotide fits the consensus sequence derived from all functionally characterized glucocorticoid response elements, shown in Figure 1b (refs 3, 4). When this sequence is altered to generate a rotationally symmetrical sequence (Fig. 1c), inducibility is maintained, suggesting that the twofold symmetry could be important for recognition by the receptor⁵. It

```

a TGTACAGGATGTTCT
  ACATGTCCTACAAGA

b GGTACANNNTGTTCT
  CCATGTNNNACAAGA

c AGAACATGATGTTCT
  TCTTGACTACAAGA

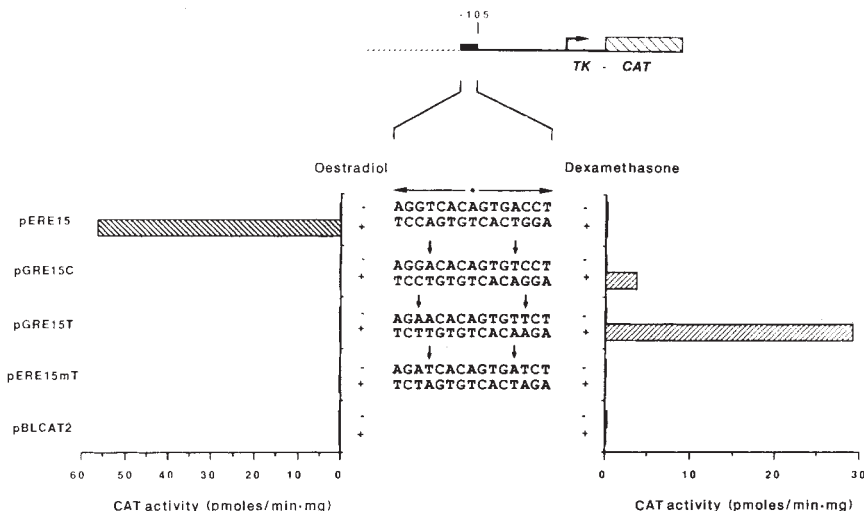
d NGGTCANNNTGACCN
  NCCAGTNNNACTGGN
  
```

Fig. 1 Comparison of the sequence of GREs with a conserved sequence in oestrogen-inducible vitellogenin genes. *a*, GRE from the rat TAT gene^{4,5}; *b*, GRE consensus sequence^{3,4}; *c*, GRE with perfect twofold symmetry⁵; *d*, conserved sequence in the 5'-flanking region of vitellogenin genes⁶. The bars indicate homologous bases between two adjacent sequences.

has been pointed out that, on the basis of sequence comparisons, several oestrogen-regulated genes from different species contain a conserved palindromic sequence in the 5'-flanking region⁶. It can be seen from Fig. 1d that this sequence element is closely related to the palindrome allowing glucocorticoid and progesterone induction (Fig. 1c). Moreover, this element is contained in the short fragments from the vitellogenin A and B genes which confer oestradiol inducibility^{8,9}.

We have investigated whether the palindromic oligonucleotide found in the vitellogenin genes could function as an oestrogen-responsive element (ERE), and also if a relationship between GREs and EREs could be demonstrated by exchange of a single or double base pair in the symmetrical motif, as suggested by the comparison in Fig. 1. To test for any correlation between the two hormone response elements, we have inserted the oligonucleotides shown in Fig. 2 into the plasmid pBLCAT2, upstream of the herpes simplex virus (HSV) thymidine kinase (TK) promoter which drives the bacterial chloramphenicol acetyltransferase (CAT) gene. After transient transfection into the human breast carcinoma cell line MCF7, which contains receptors for glucocorticoids and oestrogens¹⁰, we have determined the activity of the TK promoter either by testing for CAT enzyme activity¹¹ or with the RNase protection assay^{12,13} to measure CAT messenger RNA levels. As shown in Fig. 2, the 15-bp-long palindromic sequence (pERE15) is sufficient to

Fig. 2 Base substitutions convert an oestrogen responsive element (ERE) into a glucocorticoid responsive element (GRE). Plasmids containing the 15-bp-long sequences inserted upstream of position -105 of the HSV-TK promoter/CAT gene fusion (shown in the upper part) in pBLCAT2 (ref. 19) were used to transfect MCF7 cells. The horizontal arrows indicate the symmetry of all four oligonucleotides, the vertical arrows identify the point mutations changing equivalent bases in the palindromic halves between two sequences. The CAT activities are shown as bars, each resulting from the average of duplicate transfections. The synthetic oligonucleotides were cloned into the *Xba*I site of pBLCAT2, all DNAs were isolated from an *E. coli* strain with *dam*⁻ phenotype (GM 33), as described⁵. CAT enzymatic activity was determined after transient transfection in the absence (-) or presence (+) of either 10⁻⁷ M dexamethasone (right) or 10⁻⁸ M estradiol (left), as described⁵, except that the cells were kept in phenol-red-free medium.



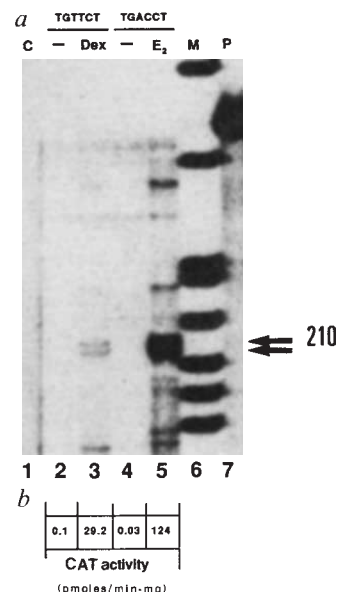
mediate a strong induction by 10^{-8} M oestradiol. In contrast, expression of pERE15 is not influenced by the synthetic glucocorticoid dexamethasone (Fig. 2). But the oligonucleotides in the two plasmids pGRE15T and pGRE15C, which contain the hexanucleotides TGTCT and TGCTCT on both sides of the palindrome, lead to induction of CAT activity by dexamethasone and not by oestradiol. pERE15mT responds to neither glucocorticoids nor oestrogens. The hormone response in the presence of dexamethasone is stronger with pGRE15T than with the variant pGRE15C. This finding is of particular interest because some GREs do contain the latter sequence, whereas the majority have a T at the fourth position of the hexanucleotide^{3,4}.

In summary, these results demonstrate firstly that a short oligonucleotide of 15 bp is sufficient to give rise to oestrogen induction, and secondly that a single base exchange in both palindromic halves of the ERE leads to a complete loss of oestrogen responsiveness and to the generation of a glucocorticoid response element.

To verify that the observed hormone stimulation of CAT expression is due to increased CAT mRNA production, we have performed an RNase protection experiment using an RNA probe specific for the TK-CAT fusion gene¹⁴. Figure 3a shows the hormone-dependent increase in CAT mRNA concentration for pERE15 (oestradiol) and pGRE15T (dexamethasone) after transient transfection of the two plasmids into MCF7 cells. The oestradiol response is stronger than the dexamethasone-dependent mRNA induction, an observation also made with parallel transfections of the two plasmids when following CAT enzyme activity (Fig. 3b and Fig. 4). In conclusion, the increase in mRNA concentration after administration of the steroids parallels the observed increase in CAT activity, and indicates that this arises from stimulation of TK promoter activity by the hormone receptors which bind to the palindromic DNA sequences upstream of the promoter region.

We have further analysed the expression of the plasmids by varying the hormone concentration (Fig. 4). Stimulation of expression of pERE15 by oestrogens (Fig. 4c) and of pGRE15T or pGRE15C (Fig. 4a and b) by glucocorticoids occurs in a concentration range that corresponds with the affinity of the receptors for these ligands¹⁵ and can be blocked by competitive inhibitors. The androgen dihydrotestosterone (DHT) gives rise to an increase of CAT expression of pERE15, but not for the two glucocorticoid-responsive plasmids (Fig. 4). The increase occurs in a concentration range in which the androgen interacts with the oestradiol receptor¹⁶. This induction by DHT is abolished by a threefold molar excess of 4-hydroxytamoxifen, a strong oestrogen antagonist, but is only slightly affected by the

Fig. 3 Hormone inducibility of CAT activity parallels increased CAT mRNA concentration. **a**, Results of an RNase protection experiment following transient transfection of MCF7 cells. Lane 1, no DNA; lanes 2 and 3, pGRE15T; lanes 4 and 5, pERE15; lane 6, pBR322 digested with *Hpa*II as a marker; lane 7, RNA probe used for the RNase protection experiment. The sequences on top indicate the differences between pGRE15T and pERE15 (compare with Fig. 2) and arrows point to the bands resulting from correctly initiated TK transcripts (210 nucleotides) (ref. 14). The cells were kept either in the absence of hormone (lanes 1, 2, 4), or in the presence of 10^{-7} M dexamethasone (lane 3) or 10^{-8} M estradiol (lane 5) respectively. Cell transfections, RNA isolation and start site mapping using an SP6 probe for the CAT gene (ref. 14) were done as described⁵. **b**, CAT activities determined in MCF7 cellular extracts after transfection in parallel and under identical conditions as for lanes 2 to 5 in **a**.



testosterone antagonist cyproterone acetate (Fig. 4d). Also the maximal response at 3×10^{-6} M DHT is the same as with 10^{-8} M oestradiol, or with both steroids together. These results show that a characteristic oestrogen response can be achieved by testosterone, which can activate the oestradiol receptor, albeit at a 1,000-fold higher concentration than the homologous ligand. The mouse mammary tumour virus (MMTV) long terminal repeat (LTR) contains an androgen-responsive element that allows testosterone induction of the MMTV LTR promoter itself¹⁷ or of the TK promoter¹⁸ at a steroid concentration typical for a hormone response mediated by the androgen receptor.

In conclusion, our results demonstrate a close sequence relationship between the oestrogen and the glucocorticoid responsive element. Related but distinct 15-bp-long sequences with palindromic structure are sufficient to confer strong inducibility on the thymidine kinase promoter by glucocorticoids and oestradiol. The symmetry of the response elements suggests a symmetrical structure for the receptors interacting with these DNA sequences; the palindromic sequence of the response

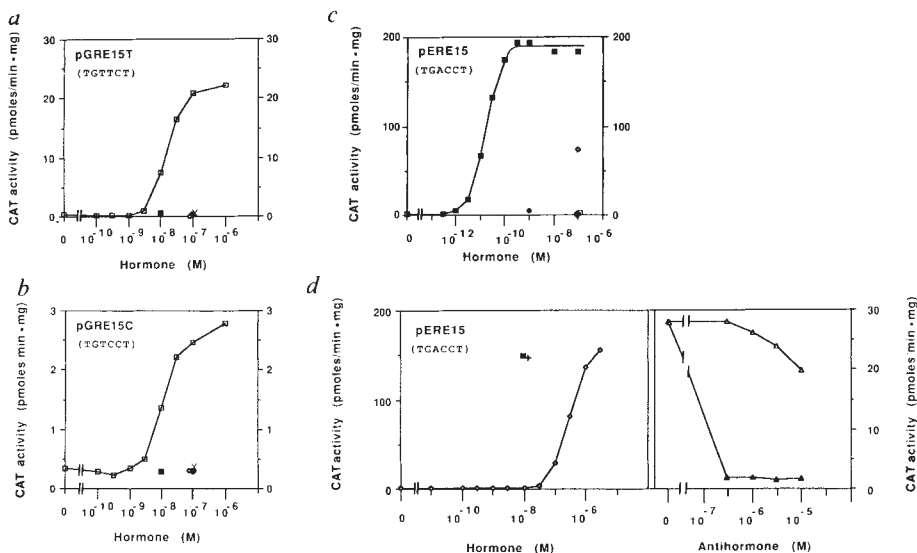


Fig. 4 Induction of CAT expression by ERE and GRE containing plasmids as a function of steroid hormone concentration. MCF7 cells were transfected with three different plasmids, as shown in the upper left hand corner of each diagram (see Fig. 2). E₂, oestradiol (■); DHT, dihydrotestosterone (◇); Dex, dexamethasone (□); R5020, (synthetic progestin) (●); OH-Tam, 4-hydroxytamoxifen (anti-oestrogen); RU486 (anti-glucocorticoid); DHT+E₂, 10^{-8} M E₂+ 10^{-7} M DHT (◊); E₂+OH-Tam, 10^{-9} M E₂+ 10^{-7} M OH-Tam (◆); Dex+RU486, 10^{-7} M Dex+ 10^{-6} M RU486 (×); DHT+CPA, 10^{-7} M DHT+cyproterone acetate (anti-testosterone, concentration as indicated) (△); DHT+OH-Tam, 10^{-7} M DHT+4 hydroxytamoxifen (concentration as indicated) (▲); concentrations of steroids alone as indicated. The transfections in Fig. 4a-c were done in parallel; the hormone administration was performed as described⁵.

element might be recognized by a dimer of two identical monomers of the oestrogen receptor. The observation that the exchange of only two bases at homologous positions within the 15-bp palindrome, altering the hexanucleotide TGACCT to TGTCTT, leads to a loss of oestrogen response and to glucocorticoid induction, corresponds with a similarity in structure (62% homology of amino-acid sequence) of the DNA-binding regions of the respective hormone receptors⁷. These findings are consistent with the probable coevolution of the DNA-binding region of steroid receptors and of the DNA sequences which they recognize.

We wish to thank Dr F. Rösl for synthesis of the oligonucleotides, Drs W. Schmid, R. Miksicek and G. Kelsey for discussion and Ms P. Di Noi for typing the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (G.K.).

Note added in proof: Similar results have been obtained by W. Wahli's group (E. Martinez *et al.* *EMBO J.*, in the press) and G. Ryffel's group (personal communication).

Received 18 May; accepted 10 September 1987.

1. Yamamoto, K. R. *A. Rev. Genet.* **19**, 209-252 (1985).
2. Ringold, G. A. *Rev. Pharmac. Tox.* **25**, 529-566 (1985).
3. Scheideit, C., Westphal, H. M., Carlson, C., Bosshard, H. & Beato, M. *DNA* **5**, 383-391 (1986).
4. Jantzen, H.-M. *et al.* *Cell* **49**, 29-38 (1987).
5. Strähle, U., Klock, G. & Schütz, G. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
6. Walker, P., Germond, J.-E., Brown-Luedi, M., Givel, F. & Wahli, W. *Nucleic Acids Res.* **12**, 8611-8626 (1984).
7. Green, S. & Chambon, P. *Nature* **324**, 615-617 (1986).
8. Klein-Hitpass, L., Schorpp, M., Wagner, U. & Ryffel, G. U. *Cell* **46**, 1053-1061 (1986).
9. Seiler-Tuyns, A. *et al.* *Nucleic Acids Res.* **14**, 8755-8770 (1986).
10. Horwitz, K. B., Zava, D. T., Thilagar, A. K., Jensen, E. M. & McGuire, W. L. *Cancer Res.* **38**, 2434-2437 (1978).
11. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
12. Zinn, K., DiMaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).
13. Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035-7056 (1984).
14. Miksicek, R. *et al.* *Cell* **46**, 283-290 (1986).
15. King, R. J. B. & Mainwaring, W. J. P. in *Steroid-Cell Interactions* (Butterworth, London, 1974).
16. Rochefort, H., Garcia, M., Vignon, F. & Westley, B. in *Steroid induced uterine proteins* (ed. Beato, M.), 171-182 (Elsevier, Amsterdam, 1980).
17. Dabre, P., Page, M. & King, R. J. B. *Molec. cell. Biol.* **6**, 2847-2854 (1986).
18. Cato, A. C. B., Henderson, D. & Ponta, H. *EMBO J.* **6**, 363-368 (1987).
19. Luckow, B. & Schütz, G. *Nucleic Acids Res.* **15**, 5490 (1987).

Different rhinovirus serotypes neutralized by anti-peptide antibodies

Joseph McCray & Gudrun Werner

Sandoz Forschungsinstitut, Brunner Strasse 59,
A-1235 Vienna, Austria

Recently, Rossman *et al.*¹ have described the three-dimensional structure of a human rhinovirus. A possible host cell surface receptor binding site was identified with a cleft on each icosahedral face. Two highly conserved amino-acid sequences found in rhino-, polio-, and foot-and-mouth disease (FMD) viruses are located near the base of this site and could be important in maintaining its topology. We have prepared site-specific antibodies^{2,3} to two synthetic peptides which include these sequences. The antibodies bind to the predicted capsid proteins of rhinovirus and neutralize ~60% of 48 rhinovirus serotypes tested. These results could provide a route to a rhinovirus vaccine effective against most of the numerous serotypes of this virus.

The two conserved viral amino-acid sequences postulated by Rossman *et al.*¹ to be structurally important in a proposed host cell receptor binding site are amino acids 151-160 (MYVPPGAPNP, using single-letter code) of capsid protein VP1 and 130-139 (AYTPPGARGP) of VP3 of the human rhinovirus serotype 14 (HRV14) (ref. 1). We have used two synthetic peptides corresponding to amino acids 147-162 (VVQAMYVPPGAPNPKEC, designated PVP1A) of VP1 and 126-141 of VP3 (KLILAYTPPGARGPQDC, designated PVP3A) of HRV14, attached to keyhole limpet haemocyanin

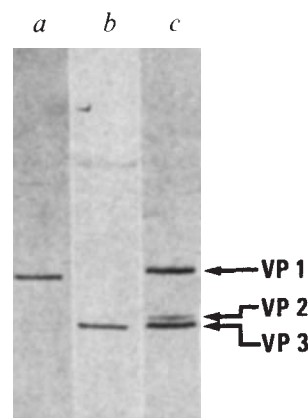


Fig. 1 Immunoblotting of rhinovirus capsid proteins with anti-peptide antibodies. *a*, Immunostaining with anti-PVP1A antibodies at a concentration of 40 µg ml⁻¹. *b*, Anti-PVP3A antibodies at 80 µg ml⁻¹. *c*, Control with an anti-HRV14 hyperimmune serum at 1:60 dilution.

Methods. HRV14 virus particles were purified as previously described¹⁵. Viral capsid proteins were separated by 5-20% SDS-polyacrylamide-gradient gel electrophoresis and blotted on to nitrocellulose sheets¹⁶. Antibodies against commercially synthesized peptides PVP1A and PVP3A (Bachem) were induced by immunization of rabbits with keyhole limpet haemocyanin conjugates¹⁷. The anti-peptide antibodies were purified from sera by affinity chromatography with columns of peptide attached to agarose. Antibody activity against the peptides was followed with ELISA (ref. 17). Visualization of immune complexes was done with a commercial horseradish-peroxidase conjugated goat anti-rabbit IgG at 1:3,000 dilution.

(KLH) by an additional cysteine residue, to raise polyclonal antibodies in rabbits. Purified antibodies were tested for binding to the capsid proteins of HRV14 and for neutralization of different rhinovirus serotypes and of other picornaviruses.

Immunoblotting (Fig. 1) demonstrated that anti-peptide antibodies bound to viral proteins as expected: antibodies to PVP1A bound strongly to VP1 and very weakly to VP3, and the opposite was found for anti-PVP3A antibodies. In enzyme-linked immunosorbent assay (ELISA) tests with the peptides (data not shown) a similar cross-reactivity was observed. This is explained by both peptides having the common sequence PPGA. A competition experiment demonstrated that the free peptides partially inhibited binding of their respective antibodies in immunoblotting (data not shown).

In ELISA tests using purified HRV14, both anti-peptide antibodies recognized whole virus (Fig. 2). Competition is shown between increasing concentrations of the respective free peptide and HRV14 for antibody binding, but is not observed for an unrelated peptide. This implies that the observed binding is site-specific (Fig. 2).

Both purified rabbit polyclonal anti-peptide antibodies neutralized HRV14 equally well (Table 1). Protection against infectivity was complete at low antibody dilutions, although titres were low compared to a rabbit antiserum against purified intact virus (Table 2). A commercial purified rabbit IgG preparation and purified rabbit antibodies against an unrelated peptide showed no neutralization. The relatively low neutralization titres agree with the frequent observation that antibodies against peptide partial sequences recognize the intact parent protein with a lower affinity compared to antibodies against the intact protein⁴. This has been noted for influenza^{3,5}, hepatitis A (ref. 6), herpes⁷ and polio viruses^{8,9}.

In an attempt to more closely define a neutralizing epitope, the affinity-purified anti-PVP1A antibodies were fractionated further on an affinity column containing bound PVP3A. About 10% of the anti-PVP1A antibodies bound to the column. These cross-reacting antibodies were found to neutralize HRV14 very

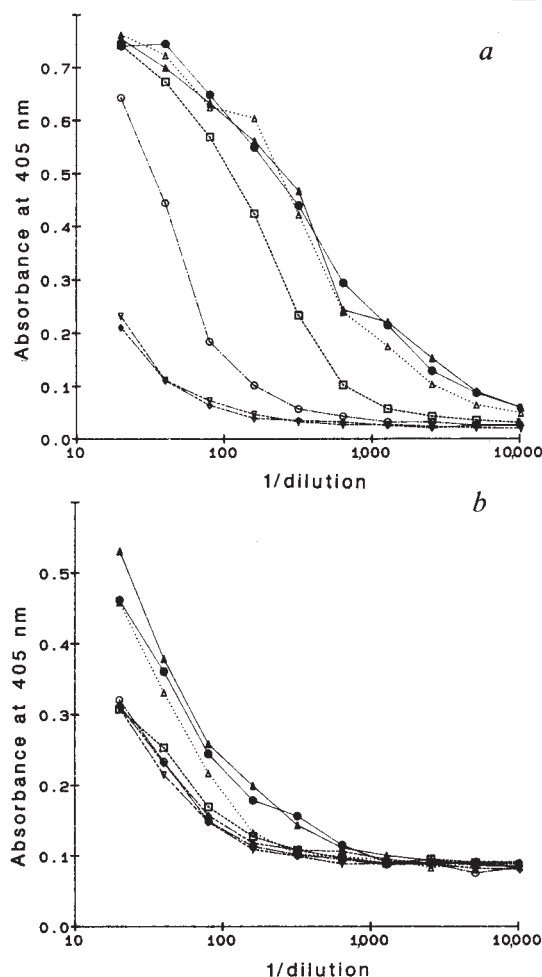


Fig. 2 ELISA to measure the binding of anti-peptide antibodies to rhinovirus 14 particles in the presence of increasing concentrations of peptide. *a*, Anti-PVP1A: $\blacktriangle$ — $\blacktriangle$, PBS+1% BSA; $\triangle$ — $\triangle$, +1 ng ml⁻¹ PVP1A; $\square$ — $\square$, +10 ng ml⁻¹; $\circ$ — $\circ$, +100 ng ml⁻¹; ∇ — ∇ , +1 μ g ml⁻¹; $\diamond$ — $\diamond$, +10 μ g ml⁻¹; $\bullet$ — $\bullet$, -10 μ g ml⁻¹ control peptide. *b*, Anti-PVP3A: $\blacktriangle$ — $\blacktriangle$, PBS+1% BSA+0.35M NaCl+0.1% Tween-20; $\triangle$ — $\triangle$, +10 ng ml⁻¹ PVP3A; $\square$ — $\square$, +100 ng ml⁻¹; $\circ$ — $\circ$, +1 μ g ml⁻¹; ∇ — ∇ , +10 μ g ml⁻¹; $\diamond$ — $\diamond$, +50 μ g ml⁻¹; $\bullet$ — $\bullet$, +50 μ g ml⁻¹ control peptide. The control peptide was a 13-amino-acid sequence from human granulocyte and macrophage colony stimulating factor with an additional C-terminal cysteine residue. Although in these experiments the virus was dried on to the plates and fixed with methanol, results are similar for virus adsorbed to plates without drying or fixation (data not shown). It is not possible to know the form of the antigen present on the surface of the ELISA plate in either test.

Methods. 50 μ l purified HRV14 (1:40 dilution in PBS at $\sim 100 \mu$ g ml⁻¹ protein, equivalent to $\sim 8 \times 10^{12}$ virus particles ml⁻¹) was added to each well of a microtitration plate. Control wells received only PBS. The plates were dried overnight at 37 °C and fixed with 50 μ l methanol per well for 5 min. The plates were dried and nonspecific binding sites blocked with 3% (w/v) bovine serum albumin (BSA) in PBS. Twofold dilutions of purified anti-PVP1A (0.58 mg ml⁻¹) and anti-PVP3A (1 mg ml⁻¹) in PBS+1% BSA (the anti-PVP3A buffer also contained 0.35M NaCl and 0.1% Tween-20) were incubated with different concentrations of either PVP1A or PVP3A in PBS+1% BSA or with a control peptide unrelated to rhinovirus. After incubation for 10 min at room temperature these samples were added to the plates containing fixed rhinovirus. The ELISA was then continued according to normal procedure. Detection of bound antibody was done with anti-rabbit immunoglobulin F(ab')₂-fragments from donkey linked to β -galactosidase (Amersham) using orthonitrophenylgalactoside as substrate. Absorbance was measured at 405 nm with an automatic microtitration plate photometer. The values shown are the average of duplicate determinations.

Table 1 Microneutralization test of HRV14 with anti-peptide antibodies

Antibody preparation	Concentration of antibodies* (μ g ml ⁻¹)	Activity in MNT (neutralizing titer)
<i>a</i> , Anti-PVP1A	18	+ (1:64)
<i>b</i> , Anti-PVP3A	12	+ (1:64)
<i>c</i> , Anti-PVP1A-PVP3A	120	—
<i>d</i> , Anti-PVP1A-PVP3A	7	+ (1:32)
<i>e</i> , Rabbit IgG	300	—

All antibody preparations except rabbit IgG were purified by affinity chromatography, using peptide Affigel columns. Peptides were coupled to Affigel by N-maleoyl- β -alanine-N-hydroxysuccinimide ester as described¹⁷. *c* and *d* are fractions from a second purification step of anti-PVP1A antibodies on a PVP3A-affinity column: *c*, contains antibodies in the flow-through and *d*, contains antibodies eluted from the column. The microneutralization test (MNT) was performed as follows: in a microtitration plate 100 μ l test antibody was serially diluted (1:2) in MEM-Eagle, supplemented with 2% FCS, 1% penicillin and streptomycin, and 40 mM MgCl₂ and preincubated for 1 h at 34 °C, after addition of 50 μ l rhinovirus, diluted in the same medium to 1×10^4 PFU ml⁻¹. 50 μ l freshly trypsinized HeLa-Ohio cells (3×10^5 cells ml⁻¹) were then added to each well and incubated at 34 °C in a humidified 5% CO₂ atmosphere until the cytopathic effect of the virus in control wells was complete (usually 2–3 days). The remaining cells were stained with crystal violet and the absorbance measured at 577 nm in an automatic ELISA-reader.

* Values indicate minimum concentrations of antibodies needed to neutralize HRV14 in MNT or the maximum concentrations tested if no neutralization was detected (*c* and *e*).

efficiently (Table 1), but antibodies in the flow-through showed almost no neutralization activity. Because PVP1A and PVP3A have only the sequence PPGA in common, these results indicate that antibodies primarily directed against this sequence can neutralize HRV14.

The high degree of conservation of the two peptide sequences in rhino, polio and FMD viruses¹ suggests that the anti-peptide antibodies could neutralize other picornaviruses as well as other rhinovirus serotypes. The neutralization by our anti-peptide antibodies of other rhinovirus serotypes was variable. Both antibodies gave similar results, and in some cases neutralization was tested with only one antibody preparation. Of the serotypes tested so far, 25% were neutralized as effectively as, or better than, HRV14; another 33% were neutralized rather less effectively than HRV14. Altogether $\sim 60\%$ of the serotypes tested were neutralized to some extent. We have also found a comparatively low degree of neutralization of a few other serotypes with an antiserum against intact purified HRV14, but, compared to the cross-reactivity seen with the anti-peptide antibodies, relatively few other serotypes were neutralized by the anti-HRV14 antiserum, and the corresponding titres were in all cases $<1\%$ of the titre against HRV14. Surprisingly, the other picornaviruses, polio (1 and 2), coxsackie B3 and echovirus 9, were not neutralized by the anti-peptide antibodies.

Recently it has been shown that the rhinoviruses can be divided into two groups recognizing different receptors¹⁰; HRV14 belongs to the larger one comprising about 80% of all serotypes¹⁰. It is interesting that no member tested from the minor group (serotypes 1a, 2, 47, and 62) was neutralized by the anti-peptide antibodies (Table 1). The sequence of HRV2 has a number of amino-acid changes in the peptide sequences tested, which might preclude recognition by the anti-peptide antibodies. Unfortunately no sequences of serotypes other than HRV14 or HRV2 are available, so we do not know if neutralization by our anti-peptide antibodies correlates with conservation of the target sequences.

Our results show that anti-peptide antibodies against sequences conserved among some picornaviruses can bind to intact virions of a large number of rhinovirus serotypes, thereby

Table 2 Neutralization of human rhinoviruses by anti-peptide antibodies

Neutralizing activity	Serotype tested
Weak (<1:8)	9†, 10, 16, 40†, 45†, 68*, 71†
Normal (1:8–1:32)	17†, 24, 26†, 28, 32†, 36, 51, 58†, 72*
Strong (>1:32)	3*, 5†, 6, 14, 23, 27, 35†, 37†, 48†, 50, 55*, 64
None	1a, 2, 4, 7, 8, 13, 15, 18, 19, 20, 21, 22, 25, 33, 38, 45, 47*, 62*, 75, 85 (polio 1, polio 2, echo 9, coxsackie B3)

Grouping of viruses was according to absorbance values (577 nm) of stained cells surviving in the virus neutralization test (as described in legend to Table 1). Reference values were from cells infected with untreated virus (blank) and from uninfected control cells (maximum values). Titres indicate dilutions of antibodies giving lysis of infected cells comparable to controls. Viruses underlined belong to a minor group of rhinoviruses using another receptor for cell penetration¹⁰.

* These serotypes were weakly positive with a rabbit hyperimmune serum against HRV14 whole virus particles. Titres ranged between 1:20 (HRV3) and 1:640 (HRV55) compared to 1:80,000 for HRV14.

† These viruses were negative with the anti-HRV14-hyperimmune serum.

neutralizing their infectivity. Therefore these sequences are probably at least partially conserved among different rhinovirus serotypes.

Although the mechanism of neutralization¹² cannot be deduced from these results, it is possible that our antipeptide antibodies block the putative cell surface receptor binding site proposed by Rossmann *et al.*¹. Both peptides include residues exposed in this site as well as exposed amino acids not actually in the site (refs 1, 13 and M. G. Rossmann, personal communication). Antibodies we presume are directed primarily against the highly conserved sequence PPGA neutralize rhinovirus (Table 1); this sequence is partially exposed in the cleft proposed as the receptor binding site, but the crystallographic data predicts that it should not be accessible to antibody molecules. It is necessary to now identify the exact contact residues in the new neutralizing epitope detected by our antipeptide antibodies.

The data reported here support the feasibility of a synthetic peptide vaccine against rhinoviruses, perhaps incorporating several short peptides necessary to ensure immunization against all the serotypes¹⁴, but the absence of an animal model makes *in vivo* testing for protection against rhinovirus infection difficult.

We are grateful to M. Pibermann, V. Weber and G. Winkler for technical assistance and to B. Rosenwirth who supplied various rhinovirus serotypes. We thank Professor M. Rossmann for discussion and communication of unpublished results.

Received 15 May; accepted 1 September 1987.

- Rossmann, M. G. *et al.* *Nature* **317**, 145–153 (1985).
- Lerner, R. A. *Adv. Immun.* **36**, 1–44 (1984).
- Shinnick, T. M., Sutcliffe, J. G., Green, N. & Lerner, R. A. *A. Rev. Microbiol.* **37**, 425–446 (1983).
- Hirayama, A., Takagaki, Y. & Karush, F. *J. Immun.* **134**, 3241–3247 (1985).
- Mueller, G. M., Shapira, M. & Arnon, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 569–573 (1982).
- Emini, E. A., Hughes, J. V., Perlow, D. S. & Boger, J., *J. Virol.* **55**, 836–839 (1985).
- Cohen, G. H. *et al.* *J. Virol.* **49**, 102–108 (1984).
- Emini, E. A., Jameson, B. A. & Wimmer, E. *J. Virol.* **52**, 719–721 (1984).
- Chow, M., Yabrov, R., Bittle, J., Hogle, J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 910–914 (1985).
- Colonna, R. J., Callahan, P. L. & Long, W. J. *J. Virol.* **57**, 7–12 (1986).
- Skern, T. *et al.* *Nucleic Acids Res.* **13**, 2111–2126 (1985).
- Dimmock, N. J., *Trends biochem. Sci.* **12**, 70–75 (1987).
- Luo, M. *et al.* *Science* **235**, 182–191 (1987).
- Chedid, L. *et al.* *Biochem. biophys. Res. Commun.* **117**, 908–915 (1983).
- Werner, G. *et al.* *J. Virol.* **57**, 1084–1093 (1986).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- Green, N. *et al.* *Cell* **28**, 477–487 (1982).

A *c-erb-A* binding site in rat growth hormone gene mediates *trans*-activation by thyroid hormone

Christopher K. Glass*, Rodrigo Franco*†, Cary Weinberger‡, Vivian R. Albert*, Ronald M. Evans‡§ & Michael G. Rosenfeld*§

* Centre for Molecular Genetics and † Department of Biology, Eukaryotic Regulatory Biology Program, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA

‡ Salk Institute, La Jolla, California 92037, USA

§ Howard Hughes Medical Institute

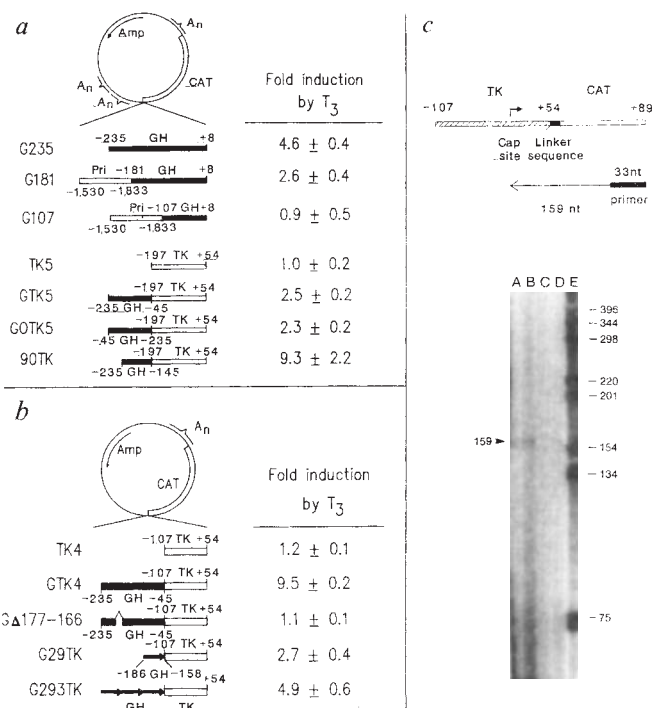
The substance 3,5,3-triiodothyronine (T_3) stimulates growth hormone gene transcription in rat pituitary tumour cells^{1–4}. This stimulation is thought to be mediated by the binding of nuclear T_3 receptors to regulatory elements 5' to the transcriptional start site^{5–8}. Understanding of the mechanism by which thyroid hormone activates gene transcription has been limited by failure to purify nuclear T_3 receptors because of their low abundance, and by the absence of defined T_3 receptor-DNA binding sites affecting T_3 regulation. Recently, human and avian *c-erb-A* gene products have been shown to bind thyroid hormone with high affinity^{9,10} and to have a molecular weight and nuclear association characteristic of the thyroid hormone receptor. In the present report, we describe the development of an avidin-biotin complex DNA-binding assay which can detect specific, high-affinity binding of rat pituitary cell T_3 receptors to the sequence 5'CAGGGACGTGACCGCA3', located 164 base pairs 5' to the transcriptional start site of the rat growth hormone gene. An oligonucleotide containing this sequence transferred T_3 regulation to the herpes simplex virus thymidine kinase promoter in transfected rat pituitary GC2 cells, and specifically bound an *in vitro* translation product of the human placental *c-erb-A* gene. The data provide supporting evidence that the human *c-erb-A* gene product mediates the transcriptional effects of T_3 and also that GC2 cell nuclear extracts contain additional factors that modify the binding of pituitary T_3 receptors to the rat growth hormone gene T_3 response element.

To identify the *cis*-active element in the growth hormone (*GH*) 5' flanking genomic sequence necessary for T_3 regulation we used a series of 5'-deleted fragments of the rat *GH* gene, fused to the bacterial chloramphenicol acetyltransferase (*CAT*) gene and transfected into rat pituitary GC2 cells (Fig. 1a). A 5' deletion to –235 base pairs (bp) from the transcriptional start (*CAP*) site transferred regulation to T_3 (4.6-fold, Fig. 1a), equivalent to constructions containing 1.7 kilobases (kb) or 307 bp of 5'-flanking rat *GH* information (data not shown), in agreement with previous studies^{5,6}. (Nucleotide positions are numbered relative to the *CAP* site, negative position numbers being 5' to it.) Deletions containing less than 235 bp of 5'-flanking *GH* information could not be assayed for T_3 induction because the levels of *CAT* expression in the absence of T_3 were not significantly above background. To overcome this problem, a rat prolactin enhancer element¹¹ was fused proximal to fragments of the rat *GH* gene containing 181 and 107 bp of 5' flanking information. The 5'-deleted fragment extending to position –181 gave 2.6-fold induction by T_3 and further deletion to position –107 from the *CAP* site abolished T_3 regulation (Fig. 1a).

To assay 3' deletions, fragments of the rat *GH* gene were fused to the herpes simplex virus thymidine kinase (*HSV tk*) promoter. A fragment of the *GH* gene extending from position –235 to position –45 from the *CAP* site produced 2.5- and 2.3-fold stimulations of *CAT* activity when fused to the *tk* promoter in native and inverse orientation, respectively (Fig. 1b). A 90-bp fragment extending from positions 235 to 145 bp

Fig. 1 Thyroid hormone responsiveness of various gene fusions containing rat *GH* 5'-flanking sequences. **a**, Responsiveness of 5' and 3' deletions of the rat *GH* gene. 5'-deleted fragments of the rat *GH* gene were fused to the CAT gene in a pSV2CAT-based vector¹⁸, in which the SV40 origin of replication and promoter were removed. These constructions were transfected into GC2 cells and assayed for responsiveness to T_3 . Because of low levels of basal expression of the nucleotide position -107 to +8 and -181 to +8 fragments, the rat prolactin enhancer (Prl)¹¹, which is not regulated by T_3 , was placed proximal to these elements. The illustrated 3'-deleted fragments were fused to a TK promoter fragment, extending from position -197 to position +54 and placed proximal to the CAT gene in the same vector. **b**, Functional analysis of the putative T_3 receptor binding site. Mutant GΔ177/166 contains a deletion of 11 bases of the T_3 receptor binding site from 177-166 bp from the CAP site. Plasmids G29TK and G293TK contain the 29-bp region of the rat *GH* gene that binds the T_3 receptor in one and three copies, respectively. The effect of T_3 was determined by dividing the percentage conversion of chloramphenicol in the presence of 10^{-9} M T_3 by the percentage conversion in the absence of T_3 . Triplicate plates were maintained in 10% fetal calf serum stripped of T_3 by ion exchange chromatography for two days before transfection with test plasmids using DEAE-dextran¹¹. The cells were treated with hormone 24 h after transfection and assayed for CAT activity after 24 h of T_3 exposure. Error limits represent the standard error of the mean. Each construction was assayed two to five times. *A_n* represents SV40 polyadenylation sites. **c**, Messenger RNA transcription initiation site analysis. The diagram indicates the 33-nucleotide primer complementary to nucleotides 67 to 89 of the *CAT* coding sequence used to determine the CAP site of transcripts of plasmids containing the TK promoter. GC cells were transfected with test plasmids and hormone treated as described for the experiments presented in panels **a** and **b**. Primer extension analysis was performed on 50 μg total RNA. Lanes A and B demonstrate the expected 159 nucleotide extension product from cells transfected with a plasmid containing the *GH* fragment extending from positions 235 to 45 from the CAP site fused to the TK promoter. Lanes C and D represent the extension product from cells transfected with a plasmid containing the TK promoter alone. No extension products were observed in these two lanes. The products shown in lanes A and C were from cells incubated in the absence of T_3 and those in lanes B and D were from cells treated with T_3 at a concentration of 10^{-9} M. Lane E shows a *Hind*III digest of pBR322, used for size calibration (in nucleotides).

Methods. Construction of *CAT* expression vectors containing 5'-flanking sequences of the rat *GH* gene from -1.7 kb to +8 bp of the CAP site has been described¹¹. Fusions containing the rat prolactin enhancer were made by excising this fragment (corresponding to rat prolactin sequence nucleotide positions 1,831-1,530) from plasmid pSSS¹¹ and inserting it in reverse orientation proximal to the 5' deletions of the growth hormone element. Fusions containing the HSV TK promoter were made by excising fragments of the *GH* enhancer from the plasmid GPO (ref. 11) and inserting them into the *Bam*HI and *Sal*I sites of a pSV2CAT-based expression vector proximal to the TK promoter at positions -197 to +54. Alternatively, the *GH* enhancer was placed proximal to the promoter at position -107 in a pUC8-based vector²⁰ by insertion into the *Bam*HI and *Sal*I polylinker sites. Site-directed mutagenesis of the *GH* enhancer element was performed by inserting the fragment from 235-45 bp from the CAP site into the *Bam*HI and *Sal*I sites of M13 mp18. A 21-base oligonucleotide was synthesized which corresponded to antisense *GH* nucleotides -188 to -157, in which nucleotides -177 to -166 were deleted and replaced by an A nucleotide. This oligonucleotide was used to delete bases -177 to -166 in the *GH* enhancer using the procedure of Kunkel²¹. *CAT* activity was determined by radioassay of methylated chloramphenicol derivatives after thin layer chromatography¹⁹. Primer extension was by the method of Elsholtz *et al.*¹².



from the *GH* CAP site transferred T_3 regulation even more efficiently than the 235-45 bp fragment, and suggests that a negative T_3 regulatory element might be located between 45 and 145 bp from the CAP site (Fig. 1a). *CAT* messenger RNA was analysed by a primer extension technique¹² to determine whether or not the T_3 -dependent stimulation of *CAT* activity which was observed in cells that had been transfected with plasmids containing the TK promoter resulted from an increase in appropriately initiated transcripts. The 235-45 bp *GH* fragment fused to the TK promoter gave an increase of about 4-fold of correctly initiated *CAT* mRNA in the presence of T_3 (Fig. 1c), a result consistent with the observed *CAT* activity measurements.

To define further the sequences needed for T_3 regulation, it was necessary to document specific binding of nuclear T_3 receptors to the growth hormone T_3 regulatory element. Because attempts to map the T_3 receptor binding site by gel shift and footprinting assays were initially unsuccessful, a new assay to detect specific binding was devised, the avidin-biotin complex DNA-binding (ABCD) assay. Double-stranded oligonucleotides were prepared containing the 5' flanking region of the *GH* gene necessary for T_3 regulation, with biotin-11-dUTP at various positions, as shown in Fig. 2a. Initially a 77-bp oligonucleotide (G209-146), containing the genomic growth hormone sequence from positions -209 to -146, was used. T_3 receptors from GC2

nuclear extracts were labelled to high specific activity with 125 I- T_3 and incubated with this biotinylated oligonucleotide. Protein-DNA complexes were precipitated from solution after the binding reaction, using streptavidin conjugated to agarose beads. It can be seen in Fig. 2b that probe G209-146 resulted in the precipitation of 6,900 c.p.m. of 125 I- T_3 activity above background. This represents the binding of about 3.2 fmol of T_3 receptor and accounts for ~40% of the total T_3 receptor present in the binding reaction. Precipitation of 125 I- T_3 -labelled receptors was probe-dependent; <15% of total precipitated 125 I- T_3 radioactivity was recovered in the absence of G209-146. Addition of a 100-fold molar excess of unlabelled T_3 reduced the precipitated 125 I- T_3 to background levels (Fig. 2b), indicating that the T_3 binding protein that was being precipitated by the probe was present in limited amounts. The equilibrium binding constant for the T_3 receptor-DNA complex was estimated to be 1.4×10^{-9} M (data not shown).

To investigate whether the precipitation of labelled T_3 receptors by G209-146 was dependent on specific rat *GH* sequences, a series of biotinylated probes were prepared that had no apparent sequence similarity to the growth hormone enhancer but which were of approximately the same length as G209-146 and contained a similar number of biotin-11-dUTP residues. As shown in Fig. 2b, the addition of 100 fmol of each

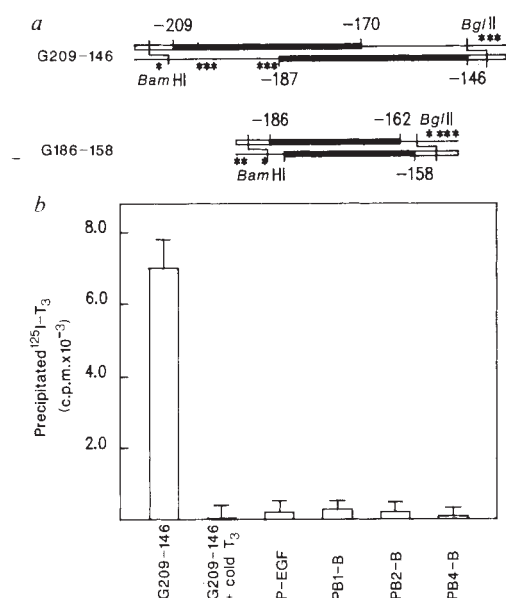


Fig. 2 Binding of T₃ receptors to oligonucleotide probes containing biotin-11-dUTP. *a*, Schematic representation of two oligonucleotide probes used to assay T₃ receptor binding to *GH* 5'-flanking sequences. Heavy lines, synthesized oligonucleotides with complementary 3' ends. Fine lines, bases incorporated by filling the 5' overhangs using the large fragment of *Escherichia coli* DNA polymerase. Asterisks, biotin-11-dUTP residues. Restriction sites for *Bam*HI and *Bgl*II are also shown. G209-146 and G186-158 contain rat *GH* enhancer sequences with the illustrated 5' and 3' boundaries. *b*, Precipitation of ^{125}I -T₃ labelled T₃ receptors from GC2 nuclear extracts by various oligonucleotide probes containing biotin-11-dUTP. P-EGF, PB1-B, PB2-B, and PB4-B are oligonucleotides of 68, 53, 55 and 58 base pairs containing 10, 11, 10 and 10 biotin-11-dUTPs respectively. These oligonucleotides contain 5'-flanking sequences of the rat prolactin gene that lack apparent homology to the rat *GH* sequence contained in G209-186. Precipitations were performed using 100 femtomole of each probe. Background, representing ^{125}I activity associated with streptavidin agarose beads in the absence of a biotinylated oligonucleotide probe, was 1,400 c.p.m. in this experiment. Results are plotted as the mean and standard error of triplicate points. The experiment is representative of six experiments examining the sequence specificity of ^{125}I -T₃ binding.

Methods. Isolated nuclei were prepared from rat GC2 cells according to the technique of Dingham *et al.*²² and salt extracted in 0.6M KCl, 10 mM Hepes, pH 7.9, 0.5 mM dithiothreitol (DTT), 0.2 mM EGTA, 20 μM ZnCl₂ for 30 min on ice. The nuclear extract was desalted by gel filtration in buffer A (50 mM KCl, 20 mM K₃PO₄, (pH 7.4) 1 mM MgCl₂, 1 mM β -mercaptoethanol and 20% glycerol) and stored at -70 °C. Assay of specific binding of T₃ to GC2 nuclear receptors was performed as described by Samuels *et al.*²³, except that T₃ binding reactions were performed in buffer A in the presence of 200 $\mu\text{g ml}^{-1}$ poly(dI-dC). To assay DNA binding, nuclear extracts were first incubated with 1 μM ^{125}I -T₃ (2,200 Ci mmol⁻¹) for 20 min at 22 °C to label the T₃ receptors to high specific activity. Aliquots (40 μl) of nuclear extract were then incubated with biotinylated probes in the presence of 200 $\mu\text{g ml}^{-1}$ poly(dI-dC) for 40 min at 22 °C. Protein-DNA complexes were precipitated by addition of 20 μl of streptavidin-agarose (BRL). Following a 10-min incubation on ice, the agarose beads were pelleted, washed three times with buffer A (1 ml) and assayed for ^{125}I activity.

probe to GC2 nuclear extracts gave no measurable precipitation of ^{125}I -T₃ above background. This indicates that the precipitation of ^{125}I -T₃ by G209-146 is dependent on the rat *GH* sequence contained in the probe.

Although early attempts to localize a T₃ receptor binding site using conventional footprinting techniques were unsuccessful, variation of buffering conditions in the ABCD binding assay suggested that a footprint might be achieved with crude nuclear

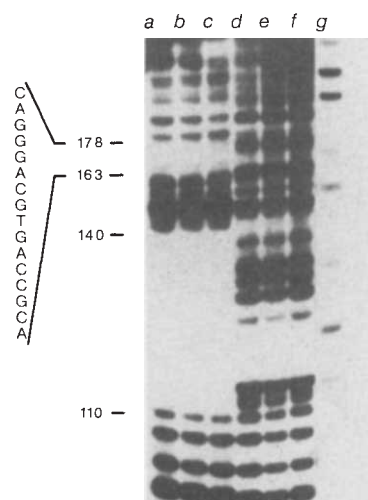


Fig. 3 DNase I footprinting of the rat *GH* enhancer element by GC2 nuclear extracts. A 16-bp protected region in the antisense strand is shown. A second footprint extending from position -110 to position -140 from the CAP site is also evident. Lanes a-c, Digestion of labelled *GH* enhancer after incubation with GC2 nuclear extract, using 24, 12 and 4 μg DNase I respectively. Lanes d-f Digestion of labelled *GH* enhancer with 24, 12 and 4 μg of DNase I in the absence of GC2 nuclear extract. Lane g, *M*, markers. The displayed sequence corresponds to that of the sense strand within the footprinted region.

Method. The antisense strand of the growth hormone enhancer was labelled with ^{32}P -ATP at its 5' end using T4 polynucleotide kinase after *Bam*HI digestion of pGPO and treatment with calf intestinal phosphatase. The enhancer fragment was released from pGPO by *Xho*I digestion and purified by agarose gel electrophoresis. Labelled *GH* enhancer fragment (1 ng, 8 fmol) was incubated with 25 μl of GC2 nuclear extracts containing 12 fmol of specific T₃ receptor-binding activity. The DNA binding reaction was carried out for 30 min at 22 °C in Buffer A. DNase digestion was for 2 min at 22 °C using the above concentrations of DNase I in a final volume of 50 μl . The reactions were stopped with 20 μl 50 mM EDTA and 1% SDS. Samples were extracted once with phenol-chloroform, ethanol precipitated, and analysed by electrophoresis on standard 10% sequencing gels.

extracts if salt and pH conditions for DNA binding were optimal (data not shown). End-labelled fragments of the *GH* enhancer were incubated with GC2 nuclear extracts and digested with DNase I. PAGE analysis of the digest under denaturing conditions (Fig. 3) gave two footprints, described previously^{13,14}; one of these is shown, together with a 16-bp protected region in the antisense strand between nucleotides -179 and -164. This sequence, which in the sense strand corresponds to 5' CAGGGACGTGACCGCA 3', is contained in the oligonucleotide probe shown to specifically bind T₃ receptors, and corresponds in position to a previously identified T₃-dependent DNase-I-hypersensitive site¹⁵. A clear footprint could not be detected in the sense strand itself, mainly because of incomplete digestion by DNase I in this G-rich region.

To examine the function of this sequence in T₃ regulation of the *GH* gene, site-directed mutagenesis was used to delete 11 bp of the footprint from the wild-type enhancer (mutant GΔ166/177). Addition of T₃ to cells transfected with mutant GΔ166/177 fused to the *tk* promoter had no effect on the amount of CAT expression, although the wild-type enhancer stimulated CAT expression ninefold (Fig. 1b). Thus, removal of the putative T₃ receptor binding region, identified by oligonucleotide and DNase I binding assays, abolished the ability of the *GH* enhancer to confer T₃ regulation to the *tk* promoter.

To demonstrate that the 16 base pairs from positions -164 to -179 constituted a functional T₃ regulatory element, a double-stranded oligonucleotide was prepared containing this sequence

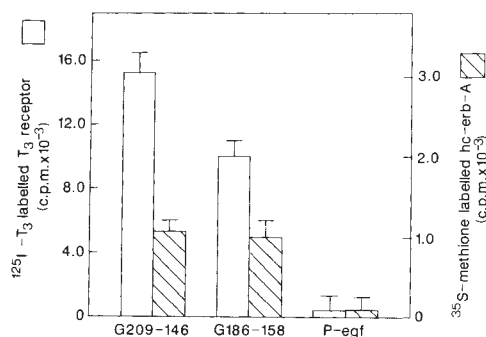


Fig. 4 Binding of rat pituitary cell T₃ receptors and an hc-erb-A *in vitro* translation product to oligonucleotides containing 64 and 29 bp of 5' flanking GH sequence. 25 µl GC2 nuclear extract were labelled with ¹²⁵I-T₃ and incubated with 100 fmol G209-146, G186-158 or P-EGF and assayed for binding as described in Fig. 2. Also, hc-erb-A *in vitro* translation product (4 µl) labelled with ³⁵S-methionine was assayed for binding to these oligonucleotides in the presence of 10 nM T₃. To prepare the hc-erb-A *in vitro* translation product, capped mRNA transcripts of an hc-erb-A cDNA were used to program translation in a rabbit reticulocyte lysate system¹⁰. Reticulocyte lysates programmed with antisense hc-erb-A mRNA had no measurable binding activity to either of the GH probes (data not shown). Results are plotted as mean and standard error of triplicate points. The experiment shown is representative of three experiments comparing the binding of GC2 nuclear T₃ receptors to the two GH probes and of four experiments examining the binding of the hc-erb-A *in vitro* translation product.

with seven and six bases flanking the 5' and 3' ends respectively. This oligonucleotide was inserted in its native orientation proximal to the *tk* promoter (construction G29TK, Fig. 1b). CAT expression was stimulated 2.9-fold by T₃ in cells transfected with this construction. Insertion of three tandem repeats of this sequence (construction G293TK) resulted in a five-fold stimulation by T₃.

The short oligonucleotide used for transfer of T₃ regulation to the *tk* promoter (G186-158) also specifically bound nuclear T₃ receptors (Fig. 4), although less efficiently than oligonucleotide G209-146. This difference could arise from an additional T₃ receptor binding site in G209-146, but this was discounted using an oligonucleotide containing the rat GH sequence from nucleotide positions -177 to -235 that failed to bind measurable amounts of T₃ receptors (data not shown). Non-biotinylated G209-146 and G186-158 were also used to compete for the binding of T₃ receptors with biotinylated G209-146: the relative affinity of T₃ receptors for G209-146 was two- to three-fold higher than G186-158 (data not shown). The apparent decrease in affinity for the shorter probe could result from a lack of bases that participate directly in the T₃ receptor binding reaction. This is unlikely, because the limits of the DNase I footprint lie within the ends of this probe. Alternatively, GC2 nuclear extracts could contain other factors that stabilize the binding of T₃ receptors to the longer probe.

These experiments demonstrated that 29 base pairs of the GH gene, containing a 16-bp footprint extending from position -164 to position -179 5' of the CAP site, bound the T₃ receptor, were necessary for T₃ regulation of the rat GH enhancer, and could transfer T₃ regulation to the *tk* promoter. To test whether human c-erb-A also binds to this element, an *in vitro* translation product was labelled with ³⁵S-methionine (ref. 10). The product migrated as a doublet of relative molecular mass (*M_r*) 48,000 (48K) and 52K on SDS gel electrophoresis and bound T₃ with a dissociation constant (*K_d*) of 5×10^{-11} Molar (data not shown). The binding of the human c-erb-A *in vitro* translation product to the G209-146 and G186-158 oligonucleotide probes containing the rat GH T₃ regulatory element is shown in Fig. 4. Both the long and short probes bound the *in vitro* translation product significantly, but probes lacking homology to the T₃ receptor binding site of the GH gene, such as P-EGF, did not. Based on the estimated specific activity of [³⁵S]methionine present in the *in vitro* transla-

tion mixture, the binding activity shown in Fig. 4 corresponds to 1–2 fmol of *erb-A* protein. Unlike GC2-cell nuclear extracts containing ¹²⁵I-T₃ receptors, the c-erb-A *in vitro* translation product was bound to the same extent by the long and short oligonucleotides. Similar data were obtained using an *in vitro* translation product of human c-erb-A labelled with ¹²⁵I-T₃ (data not shown). These results indicate that the c-erb-A gene product specifically binds to the identical T₃ regulatory sequence that is bound by T₃ receptors from GC2 nuclear extracts. They provide further evidence that the function of the c-erb-A gene product is to mediate the transcriptional effects of T₃.

Flug *et al.*⁶ recently reported that the GH sequence between nucleotide positions -210 and -181 was essential for the full stimulatory effect of T₃ in transiently transfected GC cells, and also pointed out that this region possessed limited similarity to other T₃-regulated genes. Our experiments locate the T₃ receptor-DNA binding site between 164 and 177 bp from the CAP site, and also confirm the functional importance of the sequence between positions -210 and -181 in T₃ regulation of the GH gene. This could reflect an increased affinity of the T₃ receptor for fragments of the GH enhancer which contain this upstream element *in vivo*, as we observed in the *in vitro* DNA binding studies using crude nuclear extracts as a source of T₃ receptor. That the *erb A* *in vitro* translation product binds comparably to G209-146 and G186-158 is consistent with the possibility that crude GC2 nuclear extracts contain an additional factor(s) that binds to the sequence between positions -210 and -181 and stabilizes the binding of the T₃ receptor to its cognate binding site. Cooperative interactions between eukaryotic transcription factors is well-established^{16–18}; in some cases this reflects the ability of one factor to alter the DNA-binding affinity of another. Such interactions could be important in the tissue-specific regulation of thyroid hormone action^{5,6}. The ABCD binding assay used here should be useful in addressing these questions, and it is also potentially applicable to any DNA-binding protein that can be selectively radiolabelled, either with a labelled ligand, by chemical modification or by *in vitro* translation with labelled amino acids.

We thank M. Waterman and Dr H. P. Elsholtz for discussion. This work was supported by NIH and the McKnight Foundation. C.K.G. is the recipient of a Lucille P. Markey scholarship for biomedical sciences.

Received 4 June; accepted 28 August 1987.

- Evans, R. M., Birnberg, N. C. & Rosenfeld, M. G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7659–7663 (1982).
- Diamond, D. J. & Goodman, H. M. *J. molec. Biol.* **181**, 41–62 (1985).
- Spindler, S. R., Mellon, S. H. & Baxter, J. D. *J. biol. Chem.* **257**, 11627–11632 (1982).
- Yaffee, B. M. & Samuels, H. H. *J. biol. Chem.* **259**, 6284–6291 (1984).
- Larsen, P. R., Harney, J. W. & Moore, D. D. *J. biol. Chem.* **261**, 14373–14376 (1986).
- Flug, F. *et al. J. biol. Chem.* **262**, 6373–6382 (1987).
- Casanova, J., Copp, R. P., Janocko, L. & Samuels, H. H. *J. biol. Chem.* **260**, 11744–11748 (1985).
- Crew, M. & Spindler, S. R. *J. biol. Chem.* **261**, 5018–5022 (1986).
- Sap, J. *et al. Nature* **324**, 635–640 (1986).
- Weinberger, C. *et al. Nature* **324**, 641–646 (1986).

- Nelson, C. *et al. Nature* **322**, 557–562 (1986).
- Elsholtz, H. P. *et al. Science* **234**, 1552–1557 (1986).
- West, B. L. *et al. Molec. cell. Biol.* **7**, 1193–1197 (1987).
- Catanzaro, D. F., West, B. L., Baxter, J. D. & Reudelhuber, T. L. *Molec. Endocr.* **1**, 90–96 (1987).
- Nyborg, J. K. & Spindler, S. R. *J. biol. Chem.* **261**, 5685–5688 (1986).
- Topol, J., Ruden, D. M. & Parker, C. S. *Cell*, **42**, 527–537 (1985).
- Reinberg, D., Horikoshi, M. & Roeder, R. G. *J. biol. Chem.* **262**, 3322–3330 (1987).
- McKnight, S. & Tjian, R. *Cell* **46**, 795–805 (1986).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
- Linney, E. & Donnelly, S. *Cell* **35**, 693–699 (1983).
- Kunkel, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).
- Dingham, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
- Samuels, H. H., Tsai, J. S., Casanova, J. & Stanley, F. J. *clin. Invest.* **54**, 853–865 (1974).

Genome reshuffling of the *copia* element in an inbred line of *Drosophila melanogaster*

Christian Biémont, Ammaria Aouar & Claude Arnault

Laboratoire de Biologie des Populations, Université Claude Bernard, 69622 Villeurbanne Cedex, France

Mobile genetic elements are found in the genomes of many organisms, and because of their effects on genes and their ability to induce chromosomal rearrangements they are an important source of genetic variability. Transposition rates are usually found to be low^{1,2}, estimated at around 10^{-3} per generation. Higher rates of transposition are observed, however, in crosses between certain strains of *Drosophila melanogaster* ('hybrid dysgenesis')³, which can lead to a dramatic rearrangement of many mobile elements ('transposition bursts')^{4,5}. We have studied the chromosomal distribution of *mdg-1* and *copia* mobile elements in 17 highly inbred lines of *D. melanogaster*, after 69 generations of sib-mating. Most lines show no changes, but one showed a complete reshuffling of the *copia* element. We conclude that the transpositions of the *copia* element in this line occurred rapidly in a few generations. This phenomenon, distinct from 'transposition bursts' in that only *copia* elements are involved, may account for the instability sometimes observed in inbred lines and may be important in creating genetic variability in highly homozygous populations.

Working with highly inbred lines of *D. melanogaster*, we have shown recently that, although some lines were stable for their pattern of *mdg-1* insertion sites from the 15th to the 35th generation of brother-sister mating, others showed a few gains of new insertion sites or losses of elements⁶. Hence, some inbred lines may have unstable patterns of mobile element insertions. This suggests that transposition occurs in these lines, followed by segregation over time of the new heterozygous insertion sites.

We analysed 17 inbred lines of *D. melanogaster*, each line being maintained by one sib pair every generation⁶. At generation 69, each inbred line was analysed for the number and location of the *mdg-1* and *copia* elements in its genome by *in situ* hybridization of biotinylated DNA probes on giant chromosomes (see refs 7 and 8 for the nick translation process, and ref. 9 for the *in situ* hybridization technique used). The insertion patterns on the X chromosome and on the 2L, 2R, 3L and 3R chromosome arms of the 17 inbred lines were then compared

to those observed previously at the 52nd generation¹⁰. We checked for the appearance of new sites and the disappearance of old ones. Many lines were stable between generations 52 and 69 for either *mdg-1* or *copia*, whereas other lines showed a few new insertions or losses of sites (Tables 1 and 2). However, a striking result was observed for the pattern of *copia* insertions in line 8. At generation 52, this line had 12 *copia* insertions; over the 17 following generations, it gained 11 new *copia* insertions and lost 8; hence, only four old sites were still present in larvae of the 69th generation (see Table 3). Note that this line was stable over these 17 generations for its *mdg-1* polymorphism⁶ and had the lowest copy number of *mdg-1* insertions of all the lines^{6,10}. Its *mdg-1* pattern was not changed at generation 69 (see Table 2). We also checked this line for insertions of non-*copia*-like I and P elements (involved in I-R and P-M hybrid dysgenesis) whose polymorphism was also previously analysed at the 52nd generation¹⁰. No change was observed for the I element (14 insertions). For P however, the line had lost 4 sites and then had 14 insertions at the 69th generation instead of 18 as in the 52nd (ref. 10). These changes with P perhaps resulted from residual polymorphism of this element at the 52nd generation.

The complete genome reshuffling of line 8 for the *copia* element was thus not associated with the simultaneous drastic change in pattern of the other mobile elements analysed, contrary to what is reported in hybrid dysgenesis^{4,5}. This invariance in location of some elements allows us to eliminate any hypothesis involving contamination by alien flies. To check whether the changes in *copia* insertions had been the result of one or a few old events of high rates of insertions and excisions in every generation, we analysed individually 12 larvae of line 8 at generation 69. Only two sites were polymorphic (Table 3). One was an old site seen in generation 52 (84D), the other was new (36AB). This suggests that the pattern of *copia* insertions at generation 69 was not the result of continuing movement of the element every generation, because such events should lead to extensive polymorphism of insertions between larvae, with many sites still segregating, which was not observed. Our data and the fact that the initial *copia* arrangement was not retained after the reshuffling, favour the idea of a series of *copia* transposition events in one or a few generations or in different individuals. These events probably occurred soon after generation 52, as most of the sites at generation 69 were fixed by the sib-mating system used; that there was no additional change in *copia* locations in the 70th and 71st generations reinforces this conclusion.

Table 1 Number of new and lost *copia* insertion sites in the 17 inbred lines from the 52nd to the 69th generation

Line	No. of sites at generation 52	X	2L	Chromosome arms			3R	No. of sites at generation 69
				2R	3L			
2	16	—	—	+1	—	—	—	17
3	16	—	+1	—	—	—	+1, -1	17
4	16	—	+1	—	-1	—	+1	17
5	16	—	—	—	—	—	—	16
6	17	—	—	—	—	—	—	17
7	17	—	—	—	+1	—	—	18
8	12	+3	+4, -1	-1	+3, -2	—	+1, -4	15
9	21	—	—	-1	—	—	—	20
10	22	+1	—	—	—	—	—	23
11	25	—	—	—	—	—	—	25
12	17	—	—	—	—	—	+2	19
13	18	+1	-1	—	+1	—	—	19
14	19	—	—	—	+1	—	—	20
15	18	—	—	—	—	—	—	18
16	17	+2	+1	—	—	—	—	20
17	22	+1	—	—	+1	—	—	24
18	15	—	—	—	—	—	—	15

The probe used was cDm 5002, which contains the *copia* element of 5 kilobases (kb) (refs 21 and 22). The insertion pattern shown for line 8 is the pattern of larvae with 15 insertion sites (see Table 3).

Table 2 Number of new and lost *mdg-1* insertion sites in the 17 inbred lines from the 52nd to the 69th generation

Line	No. of sites at generation 52	X	2L	Chromosome arms			No. of sites at generation 69
				2R	3L	3R	
2	16	—	—	—	—	—	16
3	16	—	—	—	—	—	16
4	14	—	—	—	—	—	14
5	15	—	-1	—	—	—	14
6	18	—	—	—	—	—	18
7	17	—	—	—	—	—	17
8	14	—	—	—	—	—	14
9	21	—	+1	—	+3	+1	26
10	16	+1	—	—	+1	+2	20
11	18	+1	—	—	—	—	19
12	15	+1	—	+1	+1	+1	19
13	15	+1	—	+1, -1	+1	—	17
14	19	—	+1	—	—	+1	21
15	15	—	—	—	—	—	15
16	16	—	—	+1	—	+1, -1	17
17	18	—	—	—	—	—	18
18	22	—	—	-1	—	—	21

The probe used consisted of a 5.6-kb fragment of *mdg-1* inserted at the *Hind*III site of plasmid pBR322 (refs 23 and 24).

The transposition bursts of various mobile elements in the course of hybrid dysgenesis in *Drosophila* have led to the hypothesis that the P factor could be an important component in the control of such events^{4,5}. This is not corroborated in our experiment in which only the *copia* element showed a high rate of transposition in line 8. Genome reshuffling of mobile elements has been reported in inbred lines in which individuals were mass-mated, suggesting a key function of the reproductive system in genome reorganization^{11,12}. No such change in the reproductive mating system can be involved in our lines because they were maintained by sib-matings every generation. The mechanism involved in genome reshuffling is unknown; indeed, because *mdg-1*, a *copia*-like element, was not affected by transposition

of the *copia* element, an overall activation of reverse transcription^{13,14} is not likely to be the explanation.

We did not detect any change in our rearing conditions which might have promoted transpositions^{15,16}. Thus we must conclude that a spontaneous high rate of transposition had occurred in our line 8. Such events, which give rise to a complete new pattern of *copia* insertion, may be responsible in some way for the instability sometimes reported in inbred lines^{17,18} and for subline differentiation¹⁹.

The evolutionary consequences of such events may be important. Indeed, a number of spontaneous and naturally occurring mutations are the result of insertions of *copia*-like or other mobile elements²⁰. Hence, a high transposition rate may provide the mobile elements with a strong mutagenic potential. This is particularly relevant for highly homozygous inbred strains. Our results show that in such strains there are mechanisms that could rapidly create new genetic variability; as a result such populations could still adapt rapidly to environmental challenges. Gross genome rearrangements involving mobile elements could thus result in a faster adaptation of genetically impoverished populations than is possible with only single base changes in structural genes. The fact that only one inbred line out of 17 shows the transposition event does not diminish its significance. We studied only four elements. Because many transposon families exist in the genome, it is likely that other elements undergo changes in pattern.

We thank C. Gautier, R. Grantham and C. Terzian for their comments, and A. Heizmann for technical assistance. This work was supported by the Centre National de la Recherche Scientifique and the Association pour la Recherche sur le Cancer.

Table 3 Insertion sites of *copia* in chromosome arms of individually analysed genomes larvae from line 8

Chromosome arm	Generation 52	Generation*
X	—	1EF (12)
	—	10DE (12)
	—	19F-20 (12)
2L	—	22A (12)
	29A	—
	35D	35D (12)
	—	36AB (6)
	—	38B (12)
	—	38C (12)
2R	57A	—
3L	—	67A (12)
	67D	—
	—	71F (12)
	—	75A (12)
3R	76B	—
	81	81 (12)
	84D	84D (5)
	85CD	—
	87A	—
	87B	—
	98B	98B (12)
	100B	—
	—	100EF (12)

* Values in parentheses give the numbers of larvae in which the insertion sites were observed.

Received 21 August; accepted 17 September 1987.

- Young, M. W. & Schwartz, H. E. *Cold Spring Harb. Symp. quant. Biol.* **45**, 629-640 (1981).
- Pierce, D. A. & Lucchesi, J. C. *Chromosoma* **82**, 471-492 (1981).
- Bregliano, J. C. & Kidwell, M. G. in *Mobile Genetic Elements* (ed. Shapiro, J. A.) (Academic, New York, 1983).
- Gerasimova, T. I., Mizrokhi, L. J. & Georgiev, G. *Nature* **309**, 714-716 (1984).
- Gerasimova, T. I., Matjunina, L. V., Mizrokhi, L. J. & Georgiev, G. *EMBO J.* **4**, 3773-3779 (1985).
- Biémont, C. & Aouar, A. *Heredity* **58**, 39-47 (1987).
- Langer-Safer, P. R., Levin, M. & Ward, D. C. *Proc. natn. Acad. Sci.* **79**, 4381-4385 (1982).
- Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. J. *molec. Biol.* **113**, 237-251 (1977).
- Biémont, C. *Chromosoma* **93**, 393-397 (1986).
- Biémont, C. & Gautier, C. *Heredity* (in the press).
- Belyaeva, E. Sp., Pasyukova, E. G., Gvozdev, V. A., Ilyin, Y. V. & Kaidanov, L. Z. *Molec. gen. Genet.* **185**, 324-328 (1982).
- Pasyukova, E. G., Belyaeva, E. Sp., Kogan, G. L., Kaidanov, L. Z. & Gvozdev, V. A. *Molec. Biol. Evol.* **3**, 299-312 (1986).
- Arkhipova, I. R. *et al. Cell* **44**, 555-563 (1986).
- Emori, Y. *et al. Nature* **315**, 773-776 (1985).
- Junakovic, N., Di Franco, C., Barsanti, P. & Palumbo, G. J. *molec. Evol.* **24**, 89-93 (1986).
- Strand, D. J. & McDonald, A. C. *Nucleic Acids Res.* **13**, 4401-4410 (1985).

17. Guillaumin, M. & Petit, C. *Annls. Génét.* 3, 9-15 (1961).
18. Petit, C. *Annls. Génét.* 6, 24-35 (1963).
19. Fitch, W. M. & Atchley, W. R. *Science* 228, 1169-1175 (1985).
20. Syvanen, M. A. *Rev. Genet.* 18, 271-293 (1984).
21. Levis, R., Dunsmuir, P. & Rubin, G. M. *Cell* 21, 581-579 (1980).
22. Dunsmuir, P., Brorén, W. J., Simon, M. A. & Rubin, G. M. *Cell* 21, 575-579 (1980).
23. Ilyin, Y. V., Chmeliauskaitė, V. G. & Georgiev, G. P. *Nucleic Acids Res.* 8, 3439-3457 (1980).
24. Tchurikov, N. A. *et al. Cold Spring Harb. Symp. quant. Biol.* 45, 655-665 (1981).

Haemonectin, a bone marrow adhesion protein specific for cells of granulocyte lineage

Alan D. Campbell*†, Michael W. Long‡
& Max S. Wicha†

* Division of Hematology/Oncology, Ann Arbor Veterans Administration Medical Center, Ann Arbor, Michigan 48109, USA

† Division of Hematology and Oncology, Department of Internal Medicine, University of Michigan, Simpson Research Institute, Ann Arbor, Michigan 48109, USA

‡ Division of Hematology and Oncology, Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, Michigan 48109, USA

There is substantial evidence that the haematopoietic microenvironment is crucial to the growth and differentiation of haematopoietic cells. This microenvironment is composed of stromal cells, soluble factors and extracellular matrix (ECM). We have shown that a complex extract of bone marrow ECM can stimulate the growth and differentiation of haematopoietic cells *in vitro*¹. Furthermore, the use of inhibitors² or stimulators³ of ECM synthesis in long-term marrow culture affects cell proliferation. On a molecular level, however, the interactions between ECM and haematopoietic cells are not well understood. We have investigated the adhesion between specific bone marrow ECM components and haematopoietic cells, and found a protein, 'haemonectin', of relative molecular mass 60,000 in bone marrow ECM which is a lineage- and organ-specific attachment molecule for cells of granulocyte lineage. This specificity distinguishes haemonectin from previously described adhesion proteins which have a wider tissue distribution and cell type specificity.

To determine whether components of bone marrow ECM promote selective attachment of specific populations of haematopoietic cells, we used murine marrow cells to compare the adhesive properties of marrow ECM and fibronectin. Fibronectin promotes the attachment of erythroid precursor cells^{4,5} and is present in the bone marrow⁶. Unfractionated murine marrow cells (C57/BL6) were plated in serum-free conditions in tissue culture dishes pre-coated with fibronectin or bone marrow ECM extract¹, and the attached cells removed after incubation by trypsin treatment and counted. Total cell attachment on the two substrates was similar, but the myeloid to erythroid ratio was 1.25 on fibronectin-coated dishes and 3.11 on bone marrow ECM-coated dishes. This suggested that bone marrow ECM contains factors which selectively promote attachment of cells of granulocytic origin.

The proteins comprising the bone marrow ECM are shown on an SDS-polyacrylamide gel in Fig. 1, lane 2. To test which might promote the selective attachment of granulocyte cell types, the marrow-derived ECM proteins were transferred to nitrocellulose after SDS polyacrylamide gel electrophoresis (SDS-PAGE) and incubated with ⁵¹Cr-labelled unfractionated bone marrow cells. Chromium-labelled marrow cells adhered strongly to a protein of relative molecular mass (M_r) 60,000 (60K) in marrow ECM (Fig. 1, lane 4). We have termed this protein haemonectin. No adhesion was seen to the protein standards (Fig. 1, lane 3). After fixation to the nitrocellulose and staining with haematoxylin, light microscopy showed that most cells had nuclear morphology consistent with a granulocyte origin.

Fig. 1 Cellular attachment to haemonectin. Cell-attachment assay shown on 10% SDS-PAGE¹²: lane 1, M_r standards visualized with silver stain, values indicated to the left; lane 2, bone marrow ECM extract visualized with silver stain. Note the intense protein band at M_r ~60K. Lanes 3 and 4, fluorograms of standards and marrow ECM after transfer to nitrocellulose, non-specific cell-binding site blockade, and incubation with ⁵¹Cr-labelled bone marrow cells. No cell attachment is seen to protein standards in lane 3 but a distinct band is seen in lane 4 which corresponds to M_r ~60K. See Table 1 for differential counts of attached and unfractionated input marrow cells.

Methods. Marrow was scraped from rabbit femurs and ECM prepared as described¹. ECM and molecular weight standards were separated by 10% SDS-PAGE¹². Lanes 1 and 2 were silver-stained¹³ and duplicate lanes were transferred to nitrocellulose¹⁴ using a modified transfer buffer (10% v/v methanol with 0.001 M SDS, 0.024 M Tris, 0.192 M glycine, pH 8.3). Efficient transfer was confirmed by Coomassie staining. After completion of electrophoretic transfer (200 mA for 16 h), the nitrocellulose was washed with 2.5% Triton X-100 in PBS for 30 min to promote renaturation of transferred proteins¹⁵, rinsed in PBS and blocked with 5% w/v evaporated milk protein ('Blotto', see ref. 16) in PBS overnight at 4°C. Bone marrow cells were flushed from femurs and tibias of C57/BL6 mice and debris removed by unit gravity sedimentation. The cells were centrifuged at 500g for 6 min and resuspended in Fisher's medium (500 µl) containing 0.1% BSA, and labelled using ⁵¹Cr at 1 µCi/10⁶ cells (New England Nuclear) for 30 min at 37°C to a specific activity of 0.03 c.p.m. per cell. The labelled cells were washed in 50 ml fresh medium, resuspended in Fisher's medium containing 0.1% albumin to a concentration of 1.2 × 10⁶ cells per ml and incubated over the nitrocellulose blot using 1.0 × 10⁶ cells cm⁻² of nitrocellulose paper for 1 h at 37°C. The blot was rinsed 5 times with PBS. After fixation with 5% v/v glutaraldehyde in PBS for 3 min, the blots were dried and visualized by fluorography (Kodak) for 48 h. For studies of cell morphology, the 60K band was located by comparison with standards and cut from the nitrocellulose strip without fixation and trypsinized. The cell suspension was then cytocentrifuged and stained.

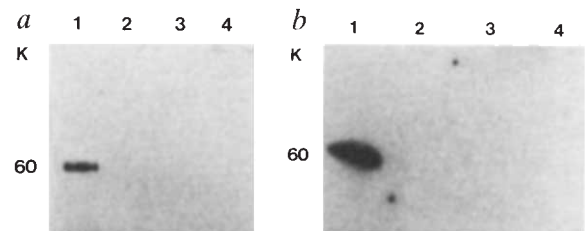
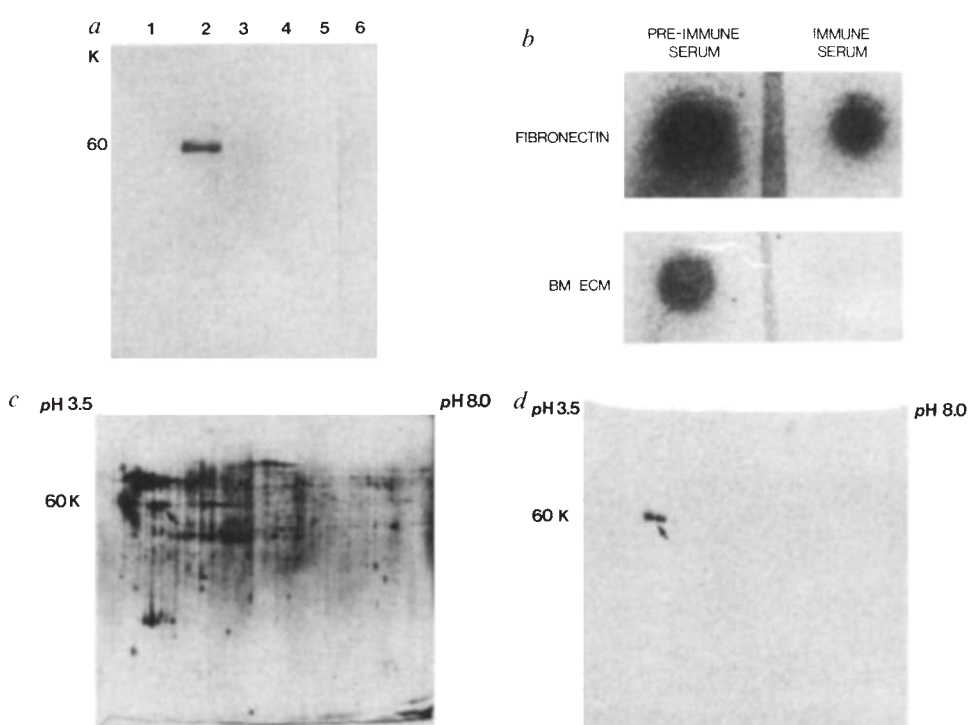


Fig. 2 Organ distribution of haemonectin. *a*, Immunoblot of bone marrow and other organ matrices using anti-haemonectin antiserum. Lane 1, bone marrow ECM; lane 2, spleen ECM; lane 3, mammary ECM; lane 4, kidney ECM. No immunological reactivity is seen in spleen, mammary or kidney ECMs. *b*, Cell attachment assay of bone marrow and non-haematopoietic tissue. Lane 1, bone marrow ECM; lane 2, spleen ECM; lane 3, mammary ECM; lane 4, kidney ECM. Attachment is seen to the 60K component of bone marrow ECM but not to components present in other ECMs.

Methods. For immunoblotting, ECMs from rabbit marrow, and rat spleen, mammary gland and kidney were prepared as described for bone marrow¹ and equal microgram quantities were separated by 10% SDS-PAGE and transferred as described. For the cell attachment assay, Western blots of equal microgram quantities of bone marrow, spleen, mammary gland and kidney ECMs were prepared, incubated with ⁵¹Cr-labelled unfractionated marrow cells, and visualized by fluorography as in Fig. 1a.

Fig. 3 Specificity of anti-haemonectin antiserum. *a*, Bone marrow ECM was compared with fibronectin, laminin, type IV collagen and vitronectin by immunoblotting using a 1:500 dilution of either pre-immune serum (lane 1) or immune serum (lanes 2–6). Lanes 1 and 2, bone marrow ECM; lane 3, fibronectin; lane 4, laminin; lane 5, type IV collagen; lane 6, vitronectin. The antibody reacts exclusively with the 60K component of bone marrow ECM. *b*, Antibody inhibition of cellular attachment. ^{51}Cr -labelled marrow cells were incubated with fibronectin or whole marrow ECM immobilized on nitrocellulose in the presence of pre-immune serum or anti-haemonectin antiserum (1:100). Adhesion is detected by fluorography. Little inhibition of adhesion to fibronectin is seen, but adhesion to marrow ECM is abolished by immune but not pre-immune serum. *c*, Two-dimensional gel electrophoresis, silver stain. *d*, Duplicate two-dimensional gel stained with anti-haemonectin by the immunoperoxidase technique as in Fig. 3*a*; anti-haemonectin recognizes only a single group of 4 closely spaced isoforms migrating at M_r 60K and pI 4.5, denoted by arrows.



Methods. Bone marrow ECM prepared as described, vitronectin (10 μg) (Calbiochem) and 50 μg each of human fibronectin (Collaborative), laminin (prepared as described¹⁷) and collagen type IV (ref. 18) were separated by 10% SDS-PAGE and transferred to nitrocellulose. The non-specific antibody-binding sites were blocked. The blot was then incubated with primary antiserum (pre-immune or immune) in a 1:500 dilution in TBS (Tris-buffered saline)-Blotto for 2 h at 20 °C. The anti-haemonectin antiserum was prepared by cutting the 60K band from a preparative gel, mixing 1:1 with complete Freund's adjuvant and injecting into the footpads of a guinea pig. The animal was boosted at week 3 and bled at week 4. The blot was then stained using a peroxidase-linked second antibody as described¹⁴. Antibody was used to block cellular attachment to fibronectin (10 μg) or whole bone marrow ECM, immobilized on nitrocellulose paper and incubated with ^{51}Cr -labelled bone marrow cells as above in the presence of either pre-immune guinea pig serum (1:100) or anti-haemonectin antiserum (1:100) for 1 h at 37 °C. The blots were then rinsed and visualized by fluorography. Two-dimensional gel electrophoresis was performed as described¹⁹.

Removal from the nitrocellulose and Romanowski staining confirmed that more than 90% of the cells had granulocyte morphology (Table 1).

To investigate the relationship of haemonectin with other ECM adhesion proteins, anti-haemonectin antiserum was raised in a guinea pig. The specificity of the antiserum was demonstrated by an immunoblot which showed reaction with the 60K component from bone marrow, but not with the defined ECM components fibronectin, laminin, type IV collagen, or vitronectin (Fig. 3*a*). The anti-haemonectin antibody also blocked cell attachment to the bone marrow ECM. Adhesion of radiolabelled bone marrow cells to fibronectin was minimally affected by anti-haemonectin antiserum but adhesion to bone marrow matrix was abrogated (Fig. 3*b*). We conclude that haemonectin is the major granulocyte attachment protein in our marrow ECM preparation and is immunologically unrelated to the defined

ECM adhesion proteins fibronectin, laminin, type IV collagen and vitronectin. We further characterized haemonectin by two-dimensional gel electrophoresis and immunoblotting of marrow ECM and found (Fig. 3*c* and *d*) that anti-haemonectin recognizes only a single group of four closely spaced isoforms migrating at M_r ~60K and pI ~4.5.

As granulopoiesis takes place primarily in bone marrow, we studied the organ distribution of haemonectin. Haemonectin could not be detected in matrix preparations from spleen, mammary gland and kidney when examined by immunoblotting (Fig. 2*a*, lanes 2–4). This organ distribution was confirmed using the cell-attachment assay (Fig. 2*b*): cell attachment occurred only at the 60K band in bone marrow ECM and there was no attachment to any components in the matrices of the other tissues. Indirect immunofluorescence with anti-haemonectin antiserum was used to study the tissue distribution of haemonectin.

Table 1 Morphology of haemonectin-adherent cells

	Immature*	Granulocytic (%) Mature†	Total	Immature‡	Erythrocytic (%) Mature§	Total
Adherent cells	38	53	91	1	4	5
Input cells	30	27	57	4	31	35

Differential counts of attached and unfractionated input marrow cells. For this morphological quantitation, cells adhering to the 60K band were removed by treatment with trypsin, cyto-centrifuged and studied under Wright staining. Cells were identified using standard morphological criteria for murine haematopoietic cells¹¹.

* Myeloblasts, progranulocytes and myelocytes.

† Metamyelocytes, bands, and PMNs.

‡ Proerythroblasts and early erythroblasts.

§ Intermediate and late erythroblasts. The remaining cells, constituting less than 5% of the total, had lymphocytic morphology, or were not identified.

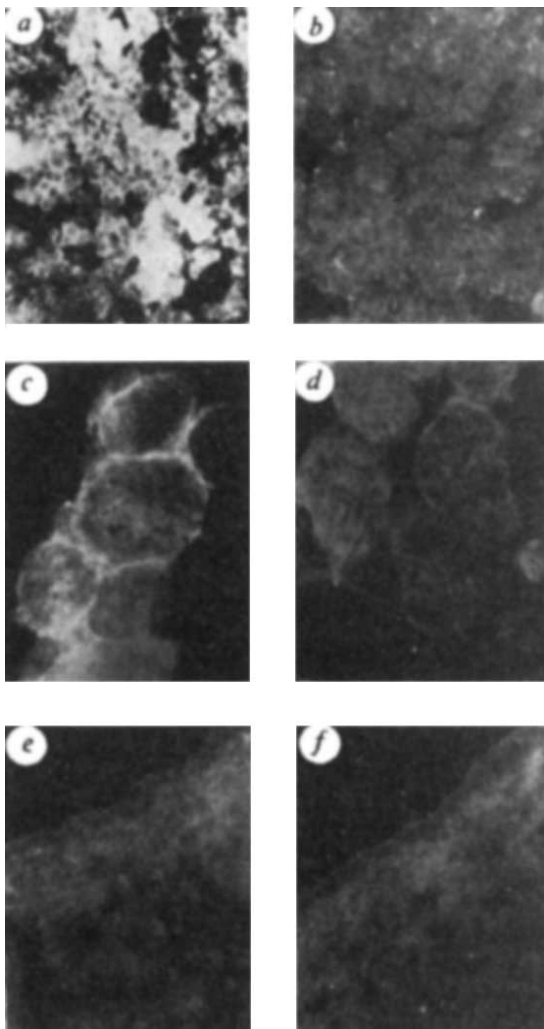


Fig. 4 Immunofluorescent localization of haemonectin *in vivo*. *a*, Bone marrow, anti-haemonectin; *b*, bone marrow, pre-immune serum; *c*, bone spicules, anti-haemonectin; *d*, bone spicules, pre-immune serum; *e*, spleen, anti-haemonectin; *f*, spleen, pre-immune serum. Note positive staining in bone marrow and endosteal surface of bone spicules.

Methods. Bone marrow was flushed from a C57/BL6 mouse femur with PBS and centrifuged at 500g for 6 min. The pellet was snap-frozen in liquid nitrogen, and stored at -70°C . For spleen tissue, C57/BL6 murine spleen was cut into small pieces and snap-frozen. Cryostat sections (8 μm thick) were mounted on polylysine-coated slides. For immunofluorescent staining the sections were fixed with 1% paraformaldehyde in PBS, pH 7.4, for 15 min at 20°C then washed with 0.1 M glycine in PBS, pH 7.4, 5 times for 5 min. The sections were then incubated with immune or pre-immune serum diluted 1:10 in PBS containing 0.1% albumin for 1 h, then with a 1:60 dilution of rhodamine isothiocyanate-conjugated goat anti-guinea pig IgG (Cooper Biomedical) for 1 h at 20°C . After rinsing, the sections were covered with PBS containing 20% glycerol, coverslipped and viewed with an epi-illuminated fluorescence microscope. Equal exposure times were used for immune and pre-immune stained sections.

tin *in vivo*: sections of murine bone marrow stained with anti-haemonectin gave specific fluorescence in the pericellular and extracellular matrix with some variation in intensity (Fig. 4*a*). In addition to staining around haematopoietic cells, bright specific fluorescence was seen on the outer (endosteal) surface of bone spicules (Fig. 4*c*). No specific fluorescence was seen in the spleen sections (Fig. 4*e*) and little or no fluorescence was seen in sections stained with preimmune sera (Fig. 4*b*, *d*, *f*).

Thus, by immunoblotting cell attachment assay and immunofluorescence, we have shown that haemonectin is found only in bone marrow.

Bone marrow ECM therefore contains an adhesive protein of $M_r \sim 60\text{K}$ and $pI \sim 4.5$, which is immunologically distinct from fibronectin, laminin, type IV collagen and vitronectin, has lineage specificity for granulocytes, and is limited to bone marrow; these features distinguish haemonectin from previously described attachment proteins, such as fibronectin, which has been implicated in the adhesion of erythroid cells^{4,5}. The lack of fibronectin-mediated attachment found in our cell attachment assay could result from lesser binding affinity between fibronectin and erythroid cells, as compared to haemonectin-granulocytic cell binding, because under the less stringent conditions of the adhesion inhibition experiment (Fig. 3*b*) fibronectin-mediated attachment was evident. Similarly, the lack of detectable cell binding to ECM components other than to a haemonectin confirms the findings of Giancotti *et al.*⁷, who found that haematopoietic cells do not attach to laminin, types I, III and IV collagens, sternal cartilage proteoglycan, or serum-spreading factor (vitronectin).

Components of the microenvironment are responsible for the granulocyte predominance of bone marrow haematopoiesis and erythroid predominance of spleen⁸. Because of its specificity for granulocyte cells and its location in bone marrow, haemonectin could explain the granulocyte predominance of haematopoiesis in this tissue. Also, the location of haemonectin could account for the homing of donor haematopoietic cells to the bone marrow in marrow transplantation. Our finding by immunofluorescence of large amounts of haemonectin along the endosteal surface of bone, an area known to contain high concentrations of stem cells^{9,10}, may bear upon this. We are currently studying the attachment of haematopoietic stem and progenitor cells to haemonectin. Peripheral blood granulocytes are only one tenth as efficient as marrow granulocyte precursors in binding to haemonectin (unpublished observations), suggesting that loss of binding to haemonectin may cause release of mature granulocytes into the circulation, a mechanism similar to that proposed for the release of mature erythrocytes from bone marrow fibronectin^{4,5}. Our experiments indicate that the haematopoietic microenvironment contains adhesion-promoting, lineage-specific ECM components in addition to the soluble factors and stromal cells previously described, and that these may be important in haematopoietic growth and differentiation.

We thank Dr Samir Hanash for technical assistance with two-dimensional gel electrophoresis and Drs Stephen Weiss, Mary Zeigler and Joanne Blum for discussion. This work was supported by the Veterans Administration, NIH, the American Cancer Society and the American Heart Association. M.W.L. is a Scholar of the Leukemia Society of America.

Received 15 June; accepted 9 September 1987.

1. Campbell, A. D., Long, M. & Wicha, M. S. *J. clin. Invest.* **75**, 2085-2090 (1985).
2. Zuckerman, K. S. *et al. J. clin. Invest.* **75**, 970-975 (1985).
3. Spooner, E., Gallagher, J. T., Krishna, F. & Dexter, T. M. *J. Cell Biol.* **96**, 510-514 (1983).
4. Patel, V. P. & Lodish, H. F. *Science* **224**, 996-998 (1984).
5. Patel, V. P. & Lodish, H. F. *J. Cell Biol.* **102**, 449-456 (1986).
6. Zuckerman, K. S. & Wicha, M. S. *Blood* **61**, 540-547 (1983).
7. Giancotti, F. G., Comoglio, P. M. & Tarone, G. *J. Cell Biol.* **103**, 429-437 (1986).
8. Wolf, N. S. & Trentin, J. J. *J. exp. med.* **127**, 205-214 (1968).
9. Lord, B. I., Testa, N. G. & Hendry, J. H. *Blood* **46**, 65-72 (1975).
10. Carsten, A. L. & Bond, V. P. *Nature* **219**, 1082-1084 (1968).
11. Chervenick, P. A., Boggs, D. R., Marsh, J. C., Cartwright, G. E. & Wintrobe, M. M. *Am. J. Physiol.* **215**, 353-360 (1968).
12. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
13. Morrissey, J. H. *Analyt. Biochem.* **117**, 307-310 (1981).
14. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4355 (1979).
15. Clarke, S. *Biochim. biophys. Acta* **670**, 195-202 (1981).
16. Johnson, D. A., Gautsch, J. A., Sportsman, J. R. & Elden, J. H. *Gene analyt. Tech.* **1**, 3-8 (1984).
17. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. *J. biol. Chem.* **252**, 1102-1106 (1977).
18. Timpl, R., Rohde, H., Robey, P. G., Rennard, S. I. & Foidart, J. M. *J. biol. Chem.* **254**, 9933-9939 (1979).
19. Hanash, S. M., Baier, L. J., McCurry, L. & Schwartz, S. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 807-811 (1986).

Analytical spectra in scientific documents

R. Manfre and L. Dumont

Integrating spectra from analytical instruments directly into scientific documents is a boon to scientific word processing. Software packages are now available for PC's and mini/mainframe computers.

THE number of scientific word processors (SWPs) available commercially has increased tremendously in the past five years. Programs are available for both mini/mainframes and personal computers (PC's), within a wide range of capabilities and prices. Scientists are becoming more aware of the usefulness of SWPs, as evidenced by the scores of articles in the literature comparing the attributes of various packages¹. Scientists are also becoming more demanding, more sophisticated users of SWPs. These machines are becoming essential in modern laboratories, and scientists are looking toward the future to predict what further capabilities should be provided in the coming years¹. The preparation of scientific documents using a computer is becoming standard.

Current scientific word processing programs perform many functions. Packages vary, but most SWPs allow researchers to write documents using full word processing functions; to use numerous fonts, such as Greek and European; and to create complex mathematical equations, molecules and reaction schemes, and tables and flow charts. Some also allow the incorporation of graphic images from other software packages, such as Lotus 1-2-3 and RS/1, into the document being prepared. But many of the images that go into scientific reports, such as spectra and chromatograms, are created outside the SWP, making integration with other instruments essential to a full-function SWP². To enable the creation of complete documents, an SWP package should provide a means of including analytical images in reports by interfacing, either directly or indirectly, with laboratory instrumentation. Few packages available today provide this capability.

Importing data

At American Cyanamid chemists have used successfully a commercially-available SWP package, ChemText, from Molecular Design Limited, to incorporate NMR spectra and HPLC chromatograms into documents created on a PC. The connection between the SWP on the PC and the NMR is direct, and involves one file-transfer step. In general, the method of transfer depends upon the hardware environment. In cases where the instrument is controlled by a PC as part of a laboratory information management system (LIMS), transfer will involve fewer steps; in other cases, it will involve more steps. The method used at American Cyanamid can be applied to any stand-alone instrument,

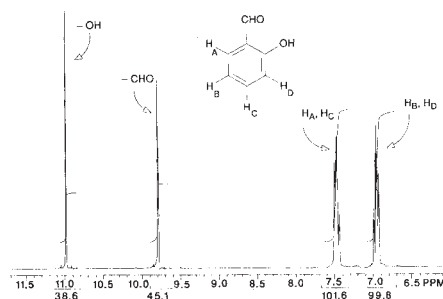


Fig. 1 An NMR file of salicylaldehyde from the Varian XL-300 Courtesy S. Alvarado.

such as an NMR, IR or HPLC, that supports Hewlett-Packard Graphic Language (HPGL) output, the language normally used to transfer information from an instrument to a Hewlett-Packard pen plotter.

The procedure to transfer information is fairly straightforward. To send information to the PC, a cable that would normally go from an RS-232 serial port on the instrument to a pen plotter is instead connected to an RS-232 serial port on the PC. Next, using a terminal emulator program, the chemist sets the configuration of the PC so that it is ready to receive data. This is done by matching communication parameters, such as baud rate, with the instrument and by setting the emulator to do an ASCII file capture. As a result, all subsequent information that is sent to the PC screen is saved to a file on the hard disk. When the chemist plots the spectrum from the instrument, the HPGL output scrolls across the screen on the PC and is saved to a file, instead of being sent to an HP plotter. The SWP package provides a utility function that converts the HPGL file format to a format it can read, allowing the spectrum to be displayed on the PC screen. The spectrum can then be manipulated on screen. It can be made larger or smaller, the pen style can be changed, and arrows, boxes or text can be added.

For example, the configuration of a Varian XL-300 NMR, which supports the HP7475 pen plotter for output, may be set to send HPGL files to a PC. To accomplish this, a 25-pin straight-through cable is run from the plotter port on the NMR to a serial port on the PC, and the terminal emulator program is started and set to perform an ASCII file capture. The spectrum can then be plotted from the NMR as if the NMR were connected to an HP pen plotter. Once captured by the PC, the HPGL file is converted to a file format understood by the SWP, and read into the SWP for inclusion in a report. The spectrum may be

enhanced on screen by inserting the chemical structure of the compound and indicating the proper chemical shift assignments (Fig. 1), and may then be inserted into a report.

A similar system may be set up to capture chromatograms from a Spectra-Physics 8800 HPLC, which stores its data on a Hewlett-Packard HP1000/3357 Laboratory Data System (LDS). Both scenarios are successful, but the SWP occasionally has difficulty incorporating very large files approaching a half megabyte.

An SWP for incorporating spectra directly from analytical instruments is also available for mini/mainframe computers. The MASS11 system by Microsystems Engineering Corporation runs on a VAX, and is an integrated, modular package of office automation software that may include word processing and several other applications, including features for creating scientific documents. A recently released optional module allows the incorporation of graphics files, including HPGL files, into documents created on a VAX. With MASS11, an image captured from an analytical instrument may be scaled, trimmed and rotated, but other images may not be superimposed over it. Text can be made to wrap around the spectrum, if so desired, but the document portion of the program does not support bit mapping, so the image cannot be seen in the document until it is printed.

The two systems discussed allow complete scientific documents, including spectra and chromatograms from laboratory instruments that support HPGL output, to be created on a computer. In the future, competitive SWPs will recognize a greater variety of plot and print files. These programs will be able to import files created by scanners, allowing scientists to include photographs and paper-based spectral and chromatographic data in their reports. They will also be able to read files of digitized video stills, such as those from scanning electron microscopy work. In the near future, a scientist will be able to incorporate virtually any image needed in a document before it is printed. □

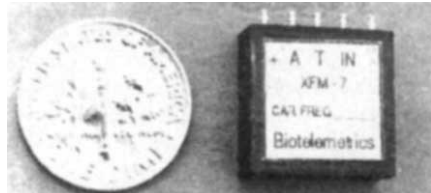
Robert Manfre is at American Cyanamid Company, PO Box 400, Princeton, New Jersey 08540 USA, and Lise Dumont is at Molecular Design Limited, 2132 Farallon Drive, San Leandro, California 94577 USA. For more information, fill in reader service number 100.

1. Marshall, J.C. *J. Chem. Inf. Comput. Sci.* **26**, 87-92 (1986).
2. Love, R.A. & Gerson, C.K. 1987 National Meeting of the American Chemical Society, Abstr. No. CINF 20.
3. Dumont, L. *Sci. comput. automat.* **3**, 16-24 (1987).
4. Dessy, R.E. *Chemometrics Intell. Lab. Sys.* **1**, 309-320 (1987).
5. Root, J.W. *Computers Sci.* **1**, 79-82 (1987).

The scientific instruments scene

Laboratories large and small can benefit from the products in this week's selection, with instruments for mass-producing microplates and washing loads of glassware, to a miniature laser.

MICROMINIATURE surgically implantable FM **telemetry transmitters** from Biotelemetry, Inc. can be used for the transmission of biologic data from unrestrained animals (*Reader Service No. 101*). The



The dime-sized Biotelemetry transmitter.

transmitters come in three types: the \$395 (US) CFM-6 transmits flow pH, temperature, pressure, or other digital high-level signals; the similarly priced CFM-8 transmits ECG, EEG, EMG or other low-level signals; and the \$255 (US) XFM-7 transmits both analog and digital data. All transmitters are encapsulated with black epoxy resin, but must be coated with beeswax or impermeable silicone prior to implantation. The battery-powered transmitters weigh no more than 1 g, and have a transmission range of up to 7 metres.

For laboratories that process large volumes of microplates for enzyme immunoassays, Skatron has an automated **microplate coater** that it says can aspirate, wash and fill 1,600 to 2,400 microplates in an 8-hour workday (*Reader Service No. 102*). The model AMS-120 machine consists of a conveyor belt, a washer, a dispenser, a microprocessor-controller and two detachable autoclavable magazines. Each magazine holds up to 120 of most common makes of microplates. During processing, plates move from one magazine, past the washer and dispenser units, and are loaded into the second magazine for storage or incubation. The microprocessor can be programmed for any required protocol, and the washer and dispenser can be configured on the transport belt to suit a variety of applications. Skatron says the \$40,000 (US) AMS-120 can dispense 100-300 µl volumes with a precision of ± 2.5 per cent C.V.

The Aristoplan microscope from Leitz is a modular universal **research microscope** that can be used for bright field, interference contrast and phase contrast techniques (*Reader Service No. 103*). The £9 000 (UK) Aristoplan has an integrated incident light fluorescent system, that allows different wavelengths to be interchanged by inserting filter modules into the changer. The mechanical stage can be rotated 90°, and the field of view

index is 28. Almost all standard objectives can be used for interference contrast work, and Wollaston prisms in modular form fit into the standard nosepiece. The Orthomat E camera system completes the Aristoplan, with its binocular head with two exit ports for camera or video systems.

Lambda Photometrics has a **miniature laser** so small that it can be held in the palm of one hand (*Reader Service No. 104*). The Gala solid-state collimated laser system uses GaAlAs laser diodes to provide power levels from 4 to 25 mW at wavelengths from 750 to 830 nm. The laser measures 50×50×107 mm, and runs on either a.c. or d.c. power. The unit requires a key to turn it on, and conforms to FDA



Lambda Photometrics' Gala laser system fits in the palm of the hand.

and BRH safety regulations. The laser has a 25 mm optic axis height and is compatible with standard optical hardware. Lambda Photometrics says the Gala laser's £512-1,132 (UK) price makes it a lower cost alternative to He:Ne lasers for industrial or laboratory applications.

A toploading microprocessor-controlled **analytical balance** is available from Fisher Scientific (*Reader Service No. 105*). The flip-up keyboard of the model XTA-100 balance can be left in place for routine simple weighings, or can be used in parts counting procedures, percentage weight-loss calculations and statistical analysis. The balance reads directly in 0.1 to 100,000 mg, 0.00001 to 3.5 Avoirdupois ounces, 0.001 to 1,543 grains and 0.0001 to 100 g. The built-in memory holds 255 weight values, permitting the operator to total them minus tare weight. The XTA-100 comes with a 108 mm pan and a draft shield, for \$1,695 (US).

Haake Buchler has a low shear-rate **viscometer**, the RV20 (*Reader Service No. 106*). The RV20 can be used manually to display 30 values of shear rate, shear



Haake Buchler's fully configured viscometer.

stress, and temperature for viscosity calculations, or can be coupled to a speed programmer and recorder for automatic plots of flow curves, yield points and thixotropy. It can also be interfaced with an IBM computer to provide visco-elastic measurements, calculations, regression models and plots. The basic \$20,000 (US) RV20 costs \$31,000 (US) with the programmer and recorder, and \$37,000 (US) with an IBM computer.

Jeol has developed a **TLC/FAB-MS system** that it says can carry out continuous, direct measurements of components separated by thin-layer chromatography with the quantitation benefits of mass spectrometry (*Reader Service No. 107*). The £20,000 (UK) system functions by introducing the TLC plate directly into the ion source of the mass spectrometer, analysing all components developed on the TLC plate by the fast-atom bombardment ionization method. The system uses either glass or aluminium TLC plates. Jeol says the system meets the high resolutions of HPTLC, and can be used in the analysis of non-volatile and thermally labile compounds.

For laboratories with lots of glassware washing to do, Lancer UK Ltd has large capacity **washing machines** that it says are capable of cleaning up after 30-50 scientists (*Reader Service No. 108*). The machines are constructed of stainless steel, and have multiple washing levels. Each microprocessor-controlled machine has a high-pressure pump coupled to a heating system with temperatures of up to 95°C. Lancer sells three large models: the £10,133 (UK) 1600E holds 320 250 ml bottles, the £6,694 (UK) 1400E holds 240, and the £5,391 (UK) 1300E holds 180. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further information, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.